Interdependent Networks from Societal Perspective: MITS(<u>M</u>ulti-context <u>I</u>nfluence <u>T</u>racking on <u>S</u>ocial Network)

RAMESH BARAL, School of Computing and Information Sciences, Florida International University
S. S. IYENGAR, School of Computing and Information Sciences, Florida International University
ASAD M. MADNI, Electrical and Computer Engineering Department, University of California Los Angeles

Abstract: The real-world system can be represented in terms of multiple complex and semantically coherent networks. The networks have some correlation among each other and complement each other's functionality. Such correlated networks are termed as interdependent networks. The notion of a smart city can be represented as an integration of several interdependent networks that can facilitate secured and efficient management of a city's assets, such as transportation, power grids, water supply channel, distributed sensor networks, societal networks, and other services. In this chapter, we introduce the societal perspective of interdependent networks, where the users' and locations' networks are exploited to track the influential user and location nodes.

The task of identifying and tracking influential nodes in the ever-growing information networks is crucial to real-world problems that require information propagation (e.g., viral marketing). The exploitation of social networks for influential node detection has been quite popular in the last decade. However, most of the studies have focused on networks with homogeneous nodes (e.g., user-user nodes), and have also ignored the impact of relevant contexts. The information networks have heterogeneous entities that are interconnected and complement each other's functionality. Hence, the classical techniques popular in modeling the spreading of epidemics in simple networks may not be efficient.

We propose a model called MITS (**<u>M</u>**ulti-context **<u>I</u>**nfluence **<u>T</u>**racking on **<u>S</u>**ocial Network) that represents the contextual exploitation of heterogeneous nodes (i.e. user-location nodes in Location-based Social Networks (LBSN)), formulates the locality-aware spatial-socio-temporal influence tracking problem using Brooks-Iyengar hybrid algorithm, and uses the geo-tagged check-in data to identify and track the locality influence. The empirical evaluation of the proposed model on two real-world datasets, using the Susceptible-Infected-Recovered (SIR) epidemic technique, coverage, and ratio of affection metrics demonstrates a significant performance gain (e.g., 10% to 85% on coverage and 14% to 39% on ratio of affection) of the proposed model against other popular techniques, such as degree centrality, betweenness centrality, closeness centrality, and PageRank.

1 INTRODUCTION

The notion of a smart city can be represented using different complex interdependent networks. The first volume of this book [1] presents the investigation of complex and interdependent networks from both theoretical and practical perspectives, including the networked control systems, graphics processing unit, smart cities, dynamic social networks, electrified transportation networks, and sustainable campus development. There are many interesting real-world examples of interdependent networks, such as power grid, transportation network, water supply channel, communication network, societal network, the Internet and the World Wide Web, points-of-interest network, phone call network, actor's collaboration network, co-authorship collaboration and citation network, genetic, metabolic and protein network, and many other relevant services [2–6]. An optimal and secured operation of resources among these interdependent networks is needed to fulfill the goal of a smart city. The flow of information among these interdependent networks can play significant role on our daily life, for instance, the real-time tracking of traffic can be propagated on social networks, helping travelers to select alternate routes, real-time tracking of weather from distributed sensors can help residents to plan their outdoor activities, tracking

Authors' addresses: Ramesh Baral, School of Computing and Information Sciences, Florida International University, Miami, FL, 33199, USA, rbara012@fiu.edu; S. S. Iyengar, School of Computing and Information Sciences, Florida International University, Miami, FL, 33199, USA, iyengar@cs.fiu.edu; Asad M. Madni, Electrical and Computer Engineering Department, University of California Los Angeles, Los Angeles, CA 90095, USA, ammadni@ee.ucla.edu.

of information flow on social networks can help to identify current trend and influential activities (e.g., trends on disasters, such as hurricane, trends on disease spread, trends on political activities, etc.), the contextual analysis of distributed sensor data can help in movement of disabled individuals [7], social network analysis (e.g., social interaction of factory workers [8], phone calls [9], communication networks on Internet [10], citation network [11]), interdependent power and transportation networks [12], metabolic networks (e.g., understanding cancer cell growth [13]), brain networks (e.g., study of brain dynamics [14–16]), community detection [17], island evolution in large epitaxial systems [18], and so forth. Despite the usefulness, the structural complexity, dynamic network and complexity, diverse nodes and connection are the major challenges behind the study of complex networks [19]. The interdependent networks are functionally correlated and the failure of a part of one network can have adverse impact on another. The exploration of their failures (e.g., blackouts resulting from cascading failures of power grid, communication systems, and financial transactions) and robustness are also of great interest [20–30] in the research community.

The graph theory [31–33] is the classical framework applied to solve many problems related to complex networks. The shortest path between nodes, clustering index, the degree centrality [34, 35], betweenness centrality [36, 37], closeness centrality [38] and k-shell decomposition methods [39], account for local or global topological network structures (see Sec. 2 for detail). The exploration of social interactions started on early 90s [40]. The exploration of student's choice on the companions [41] and social interactions between factory workers [8] are some of the early studies in social network analysis. The concept of random graphs (randomly selecting a node to form a cluster with already selected nodes) to analyze the topology of graphs [42] was used to model gene networks, ecosystems of disease and computer viruses [43, 44]. The scale-free networks (modeling the connectedness of nodes) are believed to be robust to random failures [45], and were explored for therapeutic drugs [46] and metabolic networks [46, 47]. The comprehensive studies from Boccaletti et al. [6] and Newman [48] present different phases of development of complex and interdependent networks.

One of the interesting real-world application of complex network is the social network (e.g., Facebook[1], Twitter[2], Yelp[3] etc.). The social networks can be exploited for different practical problems (e.g., friendship recommendation, tag prediction, opinion-sentiment analysis, location recommendation, item recommendation, etc.). This chapter focuses on identifying and tracking of influential nodes in LBSN. The task of influential item detection has been a popular research problem in the past decade. The influence maximization problem in a graph $G=\langle U,V \rangle$ selects a set of seed nodes in such a way that the expected spread of information (i.e., influence spreads) in the graph is maximized [49]. There are many real-world problems that exploit influence detection techniques for their business needs. As an example, in a marketing business, one can identify few influential customers and distribute promotional items to them. The strategy is to select a smallest possible subset of customers who can positively influence (e.g., via word-of mouth effect) their networks towards the consumption of the promotional items, and propagate the influence on the network (i.e. viral marketing) to an expected scale.

An interesting example is the exploitation of social networks to identify the influential candidate in U.S. presidential election. For the 2016 election, several models predicted ~80% winning chances of Hillary Clinton. Most of those studies exploited user polls, activities and posts from different social networks to capture the correlation between different parameters and to formulate the influence propagation for both candidates, and predicted the winning likelihood using their influence scores. As the correlation need not necessarily imply causation, the actual result was opposite of the predictions. This implies that those models should have been impacted by some noisy information from the social networks. For instance, for any two users, having posts in support of one candidate need not necessarily indicate their voting decision. A contextual exploitation of information on social networks is essential to model the influential parameters and to handle the noisy information.

The influential sensor measure to project the correct measure via different sensor fusion techniques [50–52] is another interesting application. The influence maximization can also be exploited on Point-of-Interest (POI)

[1] www.facebook.com [2] www.twitter.com [3] www.yelp.com

domain by exploiting the check-in preferences (e.g., likes/dislikes, tags, emoticons, reviews, stars/ratings, etc.) shared on LBSN. The sharing of positive experience can induce the linked users towards the consumption of relevant items. Intuitively, this gives us the formulation of influence tracking and influencer identification in LBSN, which can be modeled to maximize the number of potential visitors to a given POI. In this paper, we study navigation patterns of users based on LBSN data to determine influential locations (interchangeably termed as POI in this paper).

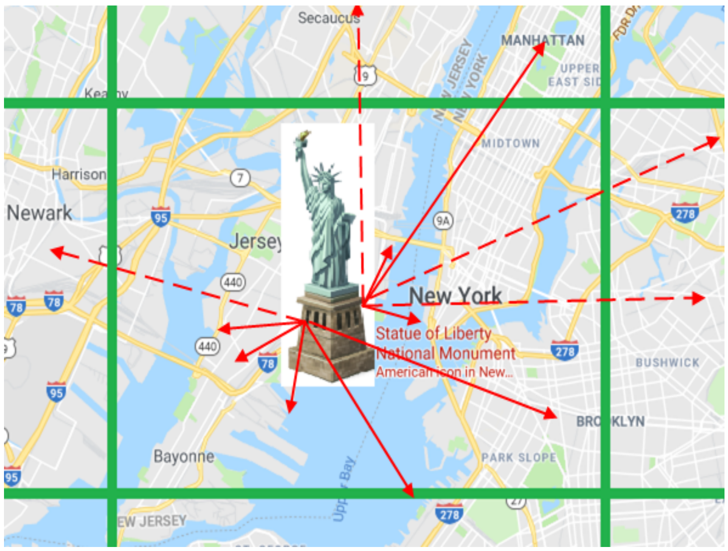

Fig. 1. POI Grid Distance Influence

Figure 1 and Figure 2 illustrates the spatial influence of a POI and user to the nearby POIs. As shown in Figure 1, the spatial influence of the *Statue of Liberty* is high to the nearby POIs. This implies most of the users who visit the *Statue of Liberty* have high chance of visiting the nearby POIs. The influence decreases with the distance, i.e. the farther POIs have less spatial influence from it. Figure 2 shows the user influence to the nearby POIs. Whenever a user has some activity on a region or a POI, the nearby POIs have high chance of being visited by the user. The category of POI, time of a day, and social relation of user also play crucial role in modeling the influence of user and POIs.

There are many interesting studies that are focused on finding influential users [53], popular events [54], or popular locations [55], but our study is focused on identifying the sets of users and POIs that have high spatial impact on other POIs. With the growing usage of smart-phones and social networks, exploitation of spatial information from mobile customers is essential in identifying the spatially influential entities, and tracking their evolution over time.

The research formulation of influence maximization was first described by Domingos and Richardson [56]. It was first formulated as a discrete optimization problem by Kempe et al. [49], who also proved it as an NP-hard problem, and proposed a greedy optimization algorithm. However, the greedy algorithm executes a Monte Carlo simulation to approximate the solution of influenced set size, and is computationally complex. The classical techniques, such as degree centrality, betweenness centrality, closeness centrality, and k-shell decomposition are quite popular in identifying influential nodes in a network. Some of the classical techniques, such as the degree centrality [34, 35], betweenness centrality [36, 37], closeness centrality [38] and k-shell decomposition

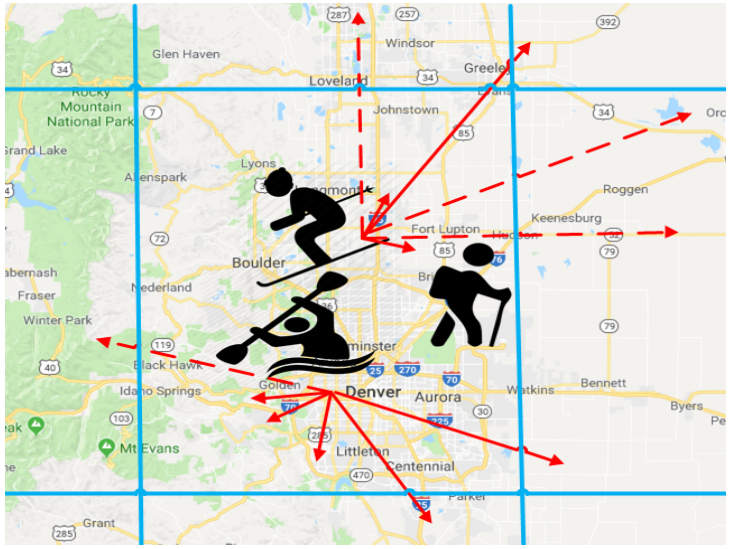

Fig. 2. User Grid Distance Influence

methods [39], account for local or global topological network structures, with some limitations. These techniques mostly shine with large networks, are unable to handle small propagation probability as they can identify only few central and overlapping nodes as the influential nodes of the network, do not handle spreading and propagation probability, and difficult to model on multi-context and heterogeneous networks.

Furthermore, the existing studies focused on location information in influence maximization only, exploited location as a simple user property, and did not analyze the contextually dynamic user mobility behaviors. In this paper, we incorporate multiple contexts (e.g., categorical, social, spatial, and temporal) of user check-ins to define the influence scores due to check-ins, and influence scores due to spatial impact. Such scores are then exploited via optimal region technique utilizing the Brooks-Iyengar algorithm [57] to find a set of users and POI influencers on location grids. The optimal region technique also helps to filter out the noisy nodes from the network. An extensive evaluation on two real-world datasets demonstrates the efficiency of our proposed model when compared to the several classical methods, such as degree centrality, closeness centrality, betweenness centrality, and PageRank-based methods. The core contributions of our study are: (1) it formulates the location promotion problem as influential user and influential POI tracking problem, (2) it presents a multi-context influence formulation by incorporating the social, temporal, categorical, and spatial attributes in the influence tracking task, and exploits an optimal region technique to find the influencing nodes in a network, and (3) it extensively evaluates the proposed model with two real-world datasets.

2 RELATED RESEARCH

The ***centrality measure*** technique is one of the classical techniques and is focused on ranking nodes in networks. A simple centrality measure is the *degree centrality* (DC) [58] of a node and is defined as the number of nearest neighbors. This technique assumes that a node with larger degree is likely to have higher influence than a node with smaller degree. However, in some cases, this method fails to identify influential nodes, since it considers only very limited information. The *closeness centrality* (CC) [38] measure of a node v is defined as the reciprocal of the sum of geodesic distances to all other nodes V in the network: $C_c(v) = \frac{1}{\sum\limits_{v' \in (V-v)} d_G(v,v')}$, where $d_G(v,v')$ is the geodesic distance between v and v'. Closeness gives a measure of how long it will spread information from a given node to other reachable nodes in the network.

The *betweenness centrality* (BC) [59] is a centrality measure of a node and is defined as the fraction of shortest paths between node pairs that pass through the node of interest. It gives a measure of the influence of a node over the information spread through the network or the expected load of a node in a transportation network. For a network G= (V,E) with n = $|V|$ nodes and m=$|E|$ edges, the *betweenness centrality* of a node v is defined as $C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where σ_{st} is the number of shortest paths between nodes s and t, and $\sigma_{st}(v)$ denotes the number of shortest paths between s and t which pass through node v. Some studies [60–63] have shown that the betweenness and closeness centrality measures can better quantify the influence of a node, but have higher computational complexity. They also claimed that the centralities based on PageRank [61] are even more relevant but more time-consuming. Though these classical techniques may work for simple networks, they have no provision for information rich networks, such as LBSN that contains multiple contexts.

The ***k-shell decomposition*** method is another popular approach and is implemented by repeatedly deleting the nodes with degree one, and then nodes with degree two, and so on. The first iteration of node deletion is repeated until all nodes' degrees are larger than one. All of these removed nodes are assigned to 1-shell. Then recursively all the nodes with degree of at most two are removed until all nodes' degrees are larger than two. The removed nodes are assigned to 2-shell. The process is repeated until all of the nodes are assigned to one of the shells [64]. In some cases, the centrality measures can have little effect on the range of the spreads and can have less impact than the strategically oriented nodes in the center of a network and with smaller degree. Some of the existing studies [39] have also shown that k-shell method can outperform the degree centrality index in many real networks.

The **Independent Cascade Model (ICM)** [49] exploits the propagation probability in a graph. Given an edge ⟨u,v⟩, and node u is active at time t, the propagation of edge ⟨u,v⟩ at time *t+1* is defined as p(u,v). The process starts with some seed nodes as active nodes and rest as inactive. The propagation is controlled using some measures, such as node degrees [49], and the process is repeated until some node gets activated. Barbieri et al. [65] extended a similar concept and incorporated topic-based information propagation. Zhu et al. [66] proposed the Gaussian-based and distance-based user mobility models, to formulate the location aware propagation probability for location promotion. Their model focused on finding seed users that can influence check-ins to a given location. They did not focus on finding the influential locations.

Wang et al. [67] proposed a community-based greed algorithm to find the influential nodes. They extended the basic independent cascade model and exploited the dynamic programming and information diffusion among the nodes to split them into small communities. Lu et al. [60] proposed a random-walk-based model called LeaderRank and identified influencers in social networks. It outperformed the PageRank [61] algorithm in identifying the influential nodes for opinion spreading and protecting from the spammers' attacks. Such ranking-based models may not be efficient for undirected networks which will degenerate the degree centrality. Chen et al. [68] proposed a local centrality measure that used the nearest and next nearest neighbors. The local centrality $C_L(v)$ of a node v was defined as: $Q(u) = \sum_{w \in \tau(u)} N(w); C_L(v) = \sum_{u \in \tau(v)} Q(u)$, where $\tau(u)$ is the set of nearest neighbors of node u and $N(w)$ is the number of the nearest and the next nearest neighbors of node w. Zhang et al. [69] used the information transfer probability between any pair of nodes and the k-medoid clustering algorithm for identifying influential nodes in complex networks with community structure.

Li et al. [70] proposed an in-degree based greedy model that focused on finding set of influencing users for a query region and the location of users. However, they did not focus on the the mobility of users. The in-degree technique also cannot capture the set of influential nodes that have smaller degrees but yet contextually relevant for the target domain. Zhu et al. [55] focused on finding the set of influential users for location promotion.

The exploitation of k-truss model from Malliaros et al. [71] also demonstrated better performance against the degree centrality measures. Recently, Wang et al. [72] proposed a multi-attribute ranking technique by exploiting the neighborhood's effect on the influence capability of a node. Wang et al. [73] proposed a fast ranking method to evaluate the influence capability of nodes using a k-shell iteration factor. They defined a relation to incorporate the degree and number of neighbors of a node to represent the influence of a node. The famous k-shell method treats nodes based on the degree at an instant and is not concerned about the original degree of the nodes and also does not address the closeness of a node to the core nodes and the location of nodes in a network.

We can see that most of the existing studies focused on identifying homogeneous influential nodes in different networks. Furthermore, exploitation of multiple contexts for heterogeneous influential node detection is less explored. Our study has following uniqueness in comparison to the existing studies: 1) it formulates the influence identification problem as a heterogeneous graph (i.e. user-POI graph) and identifies heterogeneous nodes (e.g., user and POI) as influential nodes in a network, 2) it exploits multiple contexts and defines each node with an interval of two different types of scores (i.e. score due to check-in influence and score due to spatial influence). This not only facilitates the incorporation of spatial and check-in popularity, but also handles the trade-off between the spatial and check-in influences, 3) it exploits the optimal region technique based on Brooks-Iyengar hybrid algorithm [57], and eliminates the noisy nodes that have small overlap with other nodes, and 4) it presents a simple and efficient technique for locality-based (i.e. influential nodes on different localities) influence maximization in location-based network.

3 METHODOLOGY

Many of the existing studies have defined a single composite score to identify the influential nodes in social networks. The LBSN is a special case that has many implicit factors (e.g, check-in time, location category, social relation, distance to a location, utility of a location, and so on) that can play a crucial role in identification of influential nodes. The Point-of-Interest (POI) domain can exploit such implicit factors and the explicit factors (e.g., the check-in, ratings, reviews, etc.) to define the contextual scores of a user and POI and identify the contextually influential nodes in the LBSN. Figure 3 illustrates the impact of visits and spatial impact on check-ins. The left part of the figure shows that if a POI visited at time *t* has influence on another POI, then it is most likely to be visited at time *t+1*. As an example, if the *Statue of Liberty* has visit influence on some POIs, then those POIs will be visited by the users after they visit the *Statue of Liberty*. The right part of figure shows the spatial impact. The spatial impact of POI implies that the impact of a POI to nearby places is higher than that of the farther places. For instance, the influence of the *Statue of Liberty* is higher around New York city and New Jersey city and the influence is less on farther cities.

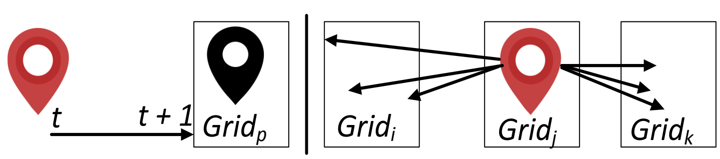

Fig. 3. Left part shows the successive check-ins for POI-checkin influence and the right part shows the spatial influence of a POI to other POIs

We divide all the locations into L uniform geographical grids (i.e., $\mathbb{L} = \{g_1, g_2, ..., g_L\}$) as in [74]. We define following two types of scores of a user and a POI on a grid (g_l) based on the check-in frequencies and the spatio-temporal influences:

(1) POI-grid visit influence (v_s^l): It is the score of a POI to a grid due to the check-in behavior of its visitors (see Figure 3 left part). The visit influence of a POI i to a grid g_l is the number of check-ins from the visitors of i to the grid g_l and is defined as:

$$y(v)_i^l = \sum_{l' \in g_l} \sum_{t=1}^{T} \sum_{t'>t}^{T} \frac{\mid V_{i,l',t} \mid}{\mid V_{l',t'} \mid}, \tag{1}$$

where the term $\mid V_{i,l',t} \mid$ is the frequency of check-ins to l' at time t and succeeding the ones to POI i, and $\mid V_{l',t'} \mid$ is the number of visits made to the POI l' at time t'.

(2) User-grid visit influence (v_s^u): It is the score of a user to a grid due to her check-in behavior. The visit influence of a user u to a grid g_l is defined in terms of the number of check-ins from her friends that succeed her check-ins to the common places, and is defined as:

$$x(v)_u^l = \sum_{l' \in g_l} \sum_{t=1}^{T} \{\alpha \times \frac{\mid V_{u,l',t} \mid}{\mid V_{l',t} \mid} + (1-\alpha) \times \sum_{t'>t}^{T} \sum_{u' \in u_f} \frac{\mid V_{u',l',t'} \mid}{\mid V_{l',t'} \mid}\}, \tag{2}$$

where $\mid V_{u,l',t} \mid$ is the frequency of check-ins made by user u to POI l' at time t, u_f is the set of friends of u, and α is a constant tuning factor.

(3) POI-grid distance influence (y_i^l): It is the influence score of a POI i to a grid g_l due to the spatial impact (see Figure 3 right part). We define a 2-D kernel density estimation to represent the influence of a POI to a grid:

$$y_i^l = \frac{1}{\sigma} K(\frac{d(i,l)}{\sigma}), \tag{3}$$

where $d(i,l)$ is the geographical distance between the POI i and the center of the grid g_l, $K(.)$ is a 2-D kernel density estimation, and σ is the standard deviation of the distances between the locations in the grid g_l. We further extend this relation to incorporate the categorical and temporal contexts, and is defined as:

$$y_i^{lt} = y_i^l + \frac{1}{\mid g_l \mid} \sum_{l' \in g_l} (C\alpha y_i^{l'} + \mathcal{T} \beta y_i^{l'}), \tag{4}$$

where C is the indicator variable to check if l' is of same category as l, $\mathcal{T}$ is the indicator variable to check if l and l' have common check-in time, and α, β are constants such that $\alpha + \beta = 1$.

(4) User-grid distance influence (x_i^{lt}): It is the influence score of a user to a grid due to spatial and social contexts. We use a 2-D kernel density estimation to represent the influence of a user at location i to a grid g_l:

$$\begin{aligned} x_i^{lt} &= \psi \times \frac{1}{\mid V_{ut} \mid} \sum_{l \in \mathbb{L}_u} \frac{\eta_u^{lt}}{\sigma} K(\frac{d(i,l)}{\sigma}) \\ &+ (1-\psi) \times \sum_{u' \in u_f} \frac{1}{\mid V_{u't} \mid} \sum_{l' \in \mathbb{L}_{u'}} \frac{\eta_u^{l't}}{\sigma} K(\frac{d(i,l')}{\sigma'}), \end{aligned} \tag{5}$$

where $\mid V_{ut} \mid$ is the number of check-ins made by user u at time t, $\mathbb{L}_u$ is the set of locations visited by user u to the grid g_l, η_u^{lt} is the number of check-ins made by user u to the grid g_l at time t, σ is the standard deviation of the distances between the locations in the grid g_l, and ψ is a constant.

These scores are normalized and each user u is represented by a pair of scores ($\min(v_s^u, x_l^u)$, $\max(v_s^u, x_l^u)$) for each grid/region g_l. A similar approach is used to define a pair of scores ($\min(v_s^l, y_l^i)$, $\max(v_s^l, y_l^i)$) for each POI l for each grid g_l. Such pair of scores are defined for each and every user and each and every POIs for a given region. We then identify the most overlapping min-max pairs. This approach provides two major benefits, first the computation cost is lowered, second the noisy nodes with very high or very low min-max scores get filtered out easily in the early steps of the process. Besides these, the model is simple and can be easily adopted to other relevant problems.

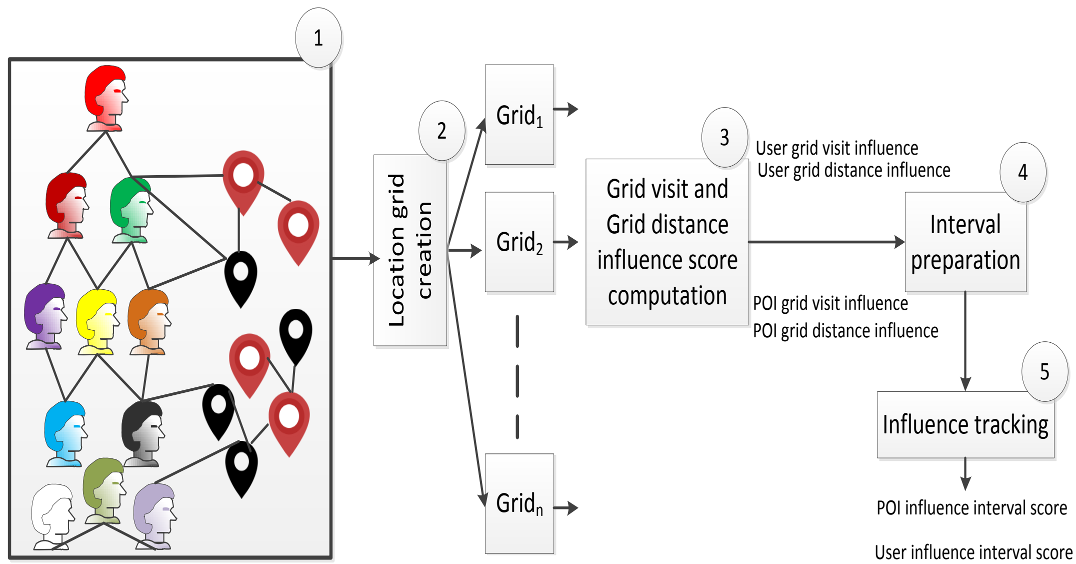

Fig. 4. High level overview of MITS

The high level overview of the proposed model is illustrated in Figure 4. The core steps of the model are as follows:

(1) The first step is to identify the target network for which the influential nodes are to be identified. The LBSN provides many contexts that can be exploited to identify the influential nodes. The LBSN can be represented as a graph that contains user and POI nodes. The user-user edges exist if the two users have social link (e.g., friendship) between them. The user-POI edges exist if the user has a check-in to the POI. The POI-POI edge exist if the POIs have same category or have distances within some threshold. There are many other contexts, such as check-in time, location category, etc.
(2) Location grid creation: It is difficult to model the contextual influence in the whole network. Inspired from [74], we divide all the locations into grids/regions of uniform areas. Each location grid contains a set of locations in the region.
(3) Grid visit and grid distance score computation: For each user and POI, we compute two different scores for each region. One of the score is due to the impact of visit and the another is due to the impact of distance. We use Equations 1- 5 to compute these scores for each user and POI nodes.
(4) Interval preparation: For each region, we need to find a measure that represents the influence score of the region. As the social network contains lot of noise, it is difficult to predict the exact score for the influential nodes, so we project an interval of score. We normalize the scores computed in the previous step and transform the scores into some buckets or intervals (e.g., {[0, 10], [11, 20], ... ,[91, 100]}).
(5) Influence tracking: We extend the Brooks-Iyengar algorithm (see Algorithm 1) to compute the interval score that represents a projection of score for a region. If the score of a node overlaps with the projected

interval, then the node is taken as an influential node. The influence tracking process is independently performed for user nodes and POI nodes in each region.

Algorithm 1 provides the detail steps of the proposed model.

Algorithm 1 Influence Tracking

1: **INPUT: (U, L, $\mathbf{G}_i$, N, k)**, **U** is the set of all users, **L** is the set of all POIs, $\mathbf{G}_i$ is the i^{th} geographical region, **N** is the number of users, **k** is the number of influencers to be identified
2: **OUTPUT:** $\mathcal{I}_i$, a list of influencer nodes for region G_i
3: get the normalized intervals of scores of all user-item nodes
4: initialize empty lists $\mathcal{I}_i$ and $\mathcal{A}_i$ to hold the influencer nodes and active nodes respectively
5: sort the user nodes by their minimum scores
6: **for** each user node u_i from step 5 **do**
7:     **if** $u_i.min$ has already been traversed but $u_i.max$ has not been traversed **then**
8:         mark this node as active, add it to $\mathcal{A}_i$
9:     **end if**
10:     /**remove the first user node whose maximum value is less than or equal to current minimum value**/
11:     **if** $\mathcal{A}_i$.size > **k and** $\mathcal{A}_i[0].max < u_i.min$ **then**
12:         remove $\mathcal{A}_i[0]$ /**remove the first non-overlapping node**/
13:     **end if**
14:     **if** $\mathcal{A}_i$.size $\geq$ **k then**
15:         add the nodes from $\mathcal{A}_i$ to $\mathcal{I}_i$, if not already present
16:     **end if**
17: **end for**
18: sort the **k** or more min-max pairs in $\mathcal{I}_i$, and use its lowest min score and highest max score to define the lower and upper influence bound of G_i
19: For the **k** min-max pairs in $\mathcal{I}_i = \langle (l_i^1, h_i^1), (l_i^2, h_i^2) \ldots (l_i^k, h_i^k) \rangle$, find a score $v_i = \frac{\sum_{n=1}^{k} \frac{(l_i^n + h_i^n)*w_i}{2}}{k}$, where w_i is the weight of the corresponding interval of min-max scores in $\mathcal{I}_i$. The score v_i represents the threshold influence score of the region, and the pair (l_i^1, h_i^k) represents the interval estimate of the influence score.

Explanation of Algorithm 1. The step 3 uses Eqns. (1- 5) to get the min-max pair scores and normalizes them. In step 4, we initialize lists to hold the active and influencer nodes. In step 5, we sort the min-max pairs. The step 6 traverses the sorted nodes. The step 7 keeps track of the traversed pairs and marks the nodes as active if their min score is already traversed and max score is not traversed. This is useful to keep track of the overlapping nodes. In step 11, the first non-overlapping pair is removed from the current active list. As we keep on adding new nodes to the active list $\mathcal{A}_i$, we check if the first node overlaps with the current node or not. There is an overlap if the pair of scores of the two nodes intersect. The higher the overlap of a pair of score, the higher the chance of it being an influential node. If there is no overlap, then the node is removed from the active nodes list. If the node was not an outlier, then it should be from an interval with high weight.

Removing the earlier entries of an interval will not be a problem because there will still be other entries from this interval with high min-max value. If the node was an outlier, then there will be only few or none entries left from this interval, hence removing the first entries with smaller max score should not be a problem. The step 14 accumulates the active nodes into the influencers list $\mathcal{I}_i$. The step 18 uses the k or more min-max pairs stored in $\mathcal{I}_i$ list to find the lowest min and highest max score to define a low-high bound for the potential influencer nodes.

This bound gives the estimated measure of the influence score interval, i.e. any node having the score overlapping with this interval is considered as a potential influencer. If we want to ensure that only top-k min-max pairs are included in the low-high bound, we can get the last k entries from the $\mathcal{I}_i$.

In step 19, we use the min-max pairs and their weight/frequency from $\mathcal{I}_i$ to find the average value of the influence score. This average value can be used as an aggregated score of influencing nodes and can be used to the scenarios that generate a single composite score. If a node's composite score falls within some range of this aggregate score, then the node can be taken as influential node. As the user and POI nodes can have different measure for the influence scores, the Algorithm 1 is independently executed for user and POI nodes.

3.1 Complexity Analysis

For each region, the sorting step takes $O(NlogN)$, where N is the number of min-max pairs. The Algorithm 1 has an additional step that removes the active nodes whose max score is less than the min score of the currently processed node. The extra step incurs a cost of removing the first item from the active list, and depending on the data structure used, the cost can range from $O(\log N)$ in the best case to $O(N)$ in worst case.

4 EVALUATION

This section presents the dataset, evaluation metrics, evaluation baselines, experimental settings, and results and discussions.

4.1 Dataset

We evaluate our proposed model MITS with two real-world datasets Weeplace [4] and Gowalla [75]. The statistics of the datasets is shown in Table 1. The datasets are well organized and have all the attributes relevant to our model.

Dataset	Check-ins	Users	Venues	Links	Location Categories
Gowalla	36,001,959	319,063	2,844,076	337,545	629
Weeplace	7,658,368	15,799	971,309	59,970	96

Table 1. Statistics of the datasets.

4.2 Evaluation Metrics

We used following evaluation metrics to evaluate the performance of the proposed model:

(1) Coverage: It measures the fraction of check-ins on a region that is induced due to the influencing user and influencing POI. We find the fraction of check-ins on test set that are done after the check-ins done on the influencing POIs. We also find the fraction of check-ins from their friends that are done after the check-ins done by the influencing users. The process is repeated for different time windows (e.g., t=2, 5, 10). The coverage of a region G_l due to a user u is defined as: $\sum_{t}^{T} \sum_{t<t'}^{T} (\sum_{\substack{l' \in G_i \\ l' \in L_u \\ u' \in u_f}} | V_{u',l',t'} | / | V_{u,l',t} |)$. The coverage of G_i due to a POI l is defined as: $\sum_{t=1}^{T} \sum_{t'>t}^{T} (\sum_{l' \in G_l} | V_{l,l',t'} | / | V_{l',t} |)$.

[4] http://www.yongliu.org/datasets/

(2) Ratio of Affection (RA): For a user u, this measures the fraction of friends that followed the check-ins of u to a region G_i, and is defined as: $\sum\limits_{\substack{l' \in G_l \\ l' \in L_u}} (\sum\limits_{u' \in u_f} \mid I(V_{u',l'}) \mid / \mid u_f \mid)$, where $I(.)$ is an indicator function. For a POI i, RA measures the fraction of POIs from a region that follow the check-ins made at i and is defined as: $\sum\limits_{l' \in G_l} (\mid I(V_{i,l'}) \mid / \mid G_l \mid)$.

(3) SIR (Susceptible, Infected, Recovered) model - In this model, for each region, a user-POI graph is built with the edges between user-user if there is a friendship relation, POI-user if there is a check-in by the user to the POI, and POI-POI edge if they are of same category and are within some threshold distance. The weight of the edges is determined by the fraction of common check-ins between two nodes. The influencing user and POI nodes are taken as the infected nodes and the spread propagation is simulated for each influencing node independently. Each influencing node undergoes 100 simulations and the average of all simulations are observed as the final measure. At a single instance of time, each infected node can infect one of the randomly selected neighbor nodes with probability (i.e. infection rate) equal to the edge weight. The infected nodes get recovered with probability λ (i.e. recovery rate) which is defined using the normalized weight of the neighbor nodes that were not infected in the original graph. The process is repeated till there is no new infected node in the graph. We observe the average infected nodes for each region and the average recovered nodes after the network gets stable.

4.3 Evaluation Baselines

We used following relevant models to evaluate the performance:

(1) Degree Centrality: It uses number of nearest neighbors of a node to compute the influence score. This model assumes that a node with larger degree has higher influence than a node with smaller degree (see 2 for detail).
(2) Closeness Centrality: It measures the reciprocal of the sum of the geodesic distances to all other nodes in the network (see 2 for detail). The node with high closeness centrality measure is taken as the highly influential node.
(3) Betweenness Centrality: It uses the fraction of shortest paths between node pairs that pass through the node of interest. The higher the fraction of shortest paths, the more influential is the node to the node pairs (see 2 for detail).
(4) PageRank [61]: The PageRank algorithm is a very popular ranking algorithm. It uses the incoming and outgoing degrees of a node in a graph to compute the rank of the node. The rank of nodes are iteratively updated until the graph converges (i.e. there is minimal or no change in the rank of the nodes).

4.4 Experimental Settings

We divided all the locations into 1,000 grids/regions with uniform area. Using the grid search technique, the parameter α in Equation 2 was checked for {0.25, 0.5, 0.75 } and the best result was found with the value of 0.5. Similarly, the values for parameters α and β was checked on the set {0.25, 0.5, 0.75 }. The best result was achieved when $\alpha = 0.25$ and $\beta = 0.75$. For each of the identified influential nodes, we ran the SIR measure for 100 times and averaged the results. The SIR measure is observed for different time steps (i.e., t=2, 5, 10, 15, 20, 25, 30).

4.5 Experimental Results and Discussion

We observe the evaluation of following models: degree centrality, betweenness centrality, closeness centrality, PageRank [61], and our proposed model MITS. We divided all the locations into 1,000 grids and observed the average values of coverage, RA, and SIR score on all of them. The coverage performance of different models

	Models	POI Coverage	User Coverage
Weeplace	DC	0.2451	0.2241
	CC	0.3626	0.2377
	BC	0.3827	0.2403
	PageRank	0.4106	0.2551
	MITS	**0.4538**	**0.2802**
Gowalla	DC	0.2590	0.2361
	CC	0.3704	0.2405
	BC	0.3883	0.2493
	PageRank	0.4216	0.2627
	MITS	**0.4637**	**0.2928**

Table 2. Coverage of different models on Weeplace and Gowalla dataset

	Models	POI RA	User RA
Weeplace	DC	0.2029	0.2035
	CC	0.2271	0.2276
	BC	0.2375	0.2337
	PageRank	0.2471	0.2401
	MITS	**0.2816**	**0.2618**
Gowalla	DC	0.2147	0.2161
	CC	0.2306	0.2306
	BC	0.2488	0.2391
	PageRank	0.2552	0.2517
	MITS	**0.2959**	**0.2704**

Table 3. Ratio of Affection (RA) of different models on Weeplace and Gowalla dataset

		Timesteps					
Dataset	**Models**	**2**	**5**	**10**	**15**	**20**	**25**
Weeplace	DC	127.26	172.38	205.17	258.71	433.51	430.25
	CC	141.81	191.50	227.20	271.92	451.61	429.01
	BC	159.20	199.29	235.39	274.99	469.28	418.30
	PageRank	165.91	218.20	242.29	291.69	485.97	410.29
	MITS	175.19	230.39	269.20	355.10	525.01	401.04
Gowalla	DC	257.46	1392.38	2281.67	2018.18	1544.51	1130.25
	CC	371.01	1471.91	2347.02	2002.20	1504.94	1115.66
	BC	402.79	1488.47	2388.83	1980.72	1483.81	1100.05
	PageRank	494.93	1527.37	2472.04	1959.37	1407.49	1063.69
	MITS	583.94	1605.70	2584.20	1920.84	1345.40	1017.47

Table 4. Average no. of infected nodes on different time steps

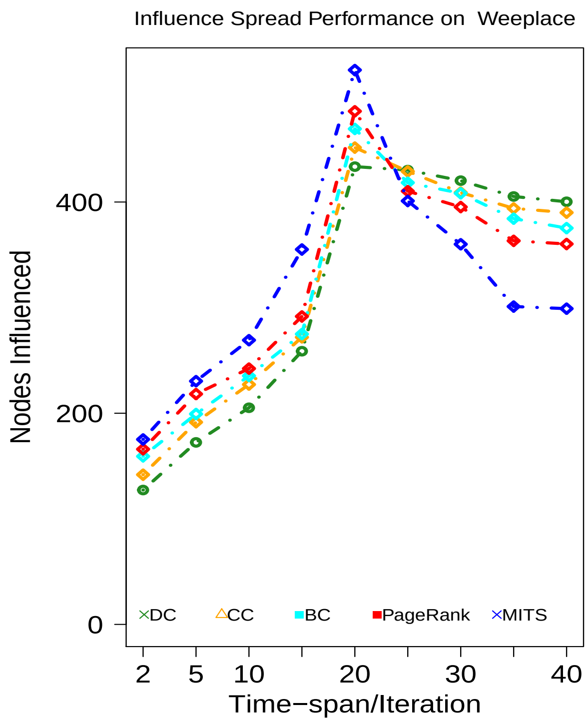

Fig. 5. Influence Spread Trend in Weeplace dataset

on Weeplace and Gowalla dataset is shown in Table 2. The coverage of DC technique is least of all the models. The PageRank-based model outperformed all other models except our proposed model. In both datasets, MITS outperformed all the other models. This implies that the influence propagation of MITS reaches to broader set of nodes when compared to the relevant models.

The ration of affection (RA) performance of different models on Weeplace and Gowalla dataset is shown in Table 3. MITS outperformed all the other models in terms of RA metric. This implies that the influential user nodes generated by MITS can impact the check-in of larger number of friend nodes in the same region. This also implies that the influential POI nodes generated by MITS can impact the higher number of successive check-ins to the POIs in the same region.

The SIR-based evaluation of different models is shown in Table 4. It shows the average number of infected nodes of all regions and all simulations for different time steps. We can see that the average number of infected nodes increases with time for all the models, until the network gets stable. The value for the DC model grows slowest in comparison to other models. The value for our model grows faster and stabilizes faster. After gaining stability, there is no infection triggered and only recovery is triggered. We can see that our model achieves faster recovery because it gets stable earlier due to the high spreading capability and the recovery also starts earlier. As the Gowalla dataset is larger, it has more social connections, has more user and POI nodes on each grid/region, hence the average number of infected and recovered nodes is higher than that of Weeplace.

The trend on SIR technique-based influence spread over time or iteration is shown in Figure 5 and Figure 6. In the Weeplace dataset (see Figure 5), all the models gain their maximum spread during time span 20. Though all of the models gain the maximum spread at this state, the number of nodes influenced varies among the models.

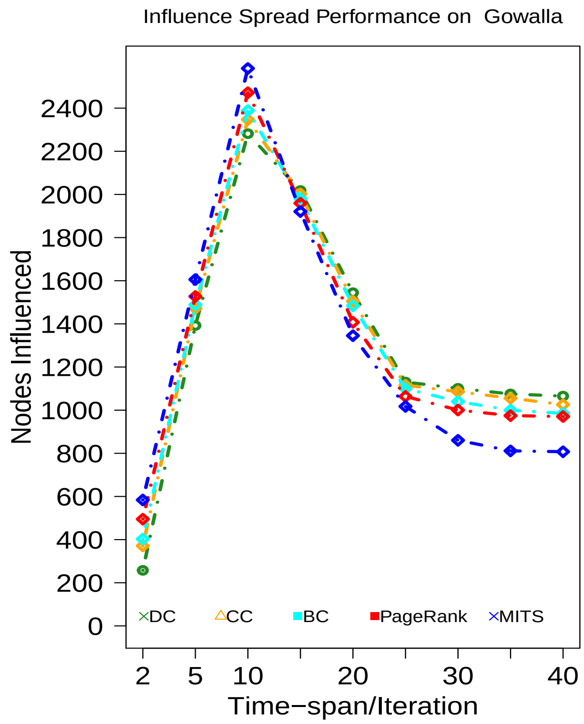

Fig. 6. Influence Spread Trend in Gowalla dataset

The DC model has least number of influenced nodes. The MITS has highest number of influenced nodes. As the time or iteration increases, the influence spread stabilizes because no new nodes are infected and the infected nodes get recovered. We can see that the rate of recovery is fastest with MITS and is slowest with the DC model. The faster recovery leads to faster stability of the network. We can see that the PageRank and MITS models are almost stable after 35 iterations but the other models still have changes in the influence spread. Using the terms from SIR technique, we can say that the spread of infection and recovery are faster in MITS, when compared to the other models. This also implies a better influence propagation.

The influence spread trend on Gowalla dataset (see Figure 6) has a spike during time span of 15 for all the models. Although all the models achieve the maximum spread at this time, the number of nodes influenced varies among the models. Similar to Weeplace dataset, the influence propagation of MITS model is better, when compared to other models. As the Gowalla dataset is larger than Weeplace dataset, the number of influenced nodes is higher for all the models. In summary, the influence propagation of MITS is better than the relevant models when evaluated on both Weeplace and Gowalla dataset.

5 CONCLUSION AND FUTURE WORK

We presented a societal perspective of complex interdependent networks by formulating the influence maximization problem on heterogeneous network (i.e. LBSN) using the optimal region maximization technique. We divided the locations into uniform grids and defined multi-context check-in based score and spati-temporal scores for each user and POI nodes. The optimal region maximization technique was independently executed to find the influencing user and influencing POI nodes. The extensive evaluation on two real-world datasets and performance metrics, such as SIR, coverage, and ratio of affection demonstrates the efficiency of our proposed model over several popular techniques. As a future work, we would like to incorporate the opinion and sentiment of users expressed in the user reviews.

ACKNOWLEDGEMENT

This research is partially supported by US Army Research Lab under the grant number W911NF-12-R-0012.

REFERENCES

[1] M. H. Amini, K. G. Boroojeni, S. Iyengar, P. M. Pardalos, F. Blaabjerg, and A. M. Madni, "Sustainable interdependent networks: From theory to application," 2018.
[2] A. Cocchia, "Smart and digital city: A systematic literature review," in *Smart city*. Springer, 2014, pp. 13–43.
[3] R. E. Hall, B. Bowerman, J. Braverman, J. Taylor, H. Todosow, and U. Von Wimmersperg, "The vision of a smart city," Brookhaven National Lab., Upton, NY (US), Tech. Rep., 2000.
[4] P. Lombardi, S. Giordano, H. Farouh, and W. Yousef, "Modelling the smart city performance," *Innovation: The European Journal of Social Science Research*, vol. 25, no. 2, pp. 137–149, 2012.
[5] H. Arasteh, V. Hosseinnezhad, V. Loia, A. Tommasetti, O. Troisi, M. Shafie-Khah, and P. Siano, "Iot-based smart cities: a survey," in *Environment and Electrical Engineering (EEEIC), 2016 IEEE 16th International Conference on*. IEEE, 2016, pp. 1–6.
[6] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
[7] P. Gharani and H. A. Karimi, "Context-aware obstacle detection for navigation by visually impaired," *Image and Vision Computing*, vol. 64, pp. 103–115, 2017.
[8] J. Scott and P. J. Carrington, *The SAGE handbook of social network analysis*. SAGE publications, 2011.
[9] W. Aiello, F. Chung, and L. Lu, "Random evolution in massive graphs," in *Handbook of massive data sets*. Springer, 2002, pp. 97–122.
[10] G. Caldarelli, F. Coccetti, and P. De Los Rios, "Preferential exchange: strengthening connections in complex networks," *Physical Review E*, vol. 70, no. 2, p. 027102, 2004.
[11] D. J. D. S. Price, "Networks of scientific papers," *Science*, pp. 510–515, 1965.
[12] M. H. Amini and O. Karabasoglu, "Optimal operation of interdependent power systems and electrified transportation networks," *Energies*, vol. 11, no. 1, p. 196, 2018.
[13] B. Vogelstein, D. Lane, and A. J. Levine, "Surfing the p53 network," *Nature*, vol. 408, no. 6810, p. 307, 2000.
[14] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, "Chaos in random neural networks," *Physical review letters*, vol. 61, no. 3, p. 259, 1988.
[15] A. A. Minai and W. B. Levy, "The dynamics of sparse random networks," *Biological Cybernetics*, vol. 70, no. 2, pp. 177–187, 1993.
[16] C. F. Mari, "Random networks of spiking neurons: instability in the xenopus tadpole moto-neural pattern," *Physical review letters*, vol. 85, no. 1, p. 210, 2000.
[17] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," *Physical review E*, vol. 72, no. 2, p. 027104, 2005.
[18] P. Mistani, A. Guittet, D. Bochkov, J. Schneider, D. Margetis, C. Ratsch, and F. Gibou, "The island dynamics model on parallel quadtree grids," 2017.
[19] S. H. Strogatz, "Exploring complex networks," *nature*, vol. 410, no. 6825, p. 268, 2001.
[20] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, p. 1025, 2010.
[21] R. Parshani, S. V. Buldyrev, and S. Havlin, "Interdependent networks: Reducing the coupling strength leads to a change from a first to second order percolation transition," *Physical review letters*, vol. 105, no. 4, p. 048701, 2010.
[22] S. V. Buldyrev, N. W. Shere, and G. A. Cwilich, "Interdependent networks with identical degrees of mutually dependent nodes," *Physical Review E*, vol. 83, no. 1, p. 016112, 2011.
[23] A. Phadke and J. S. Thorp, "Expose hidden failures to prevent cascading outages [in power systems]," *IEEE Computer Applications in Power*, vol. 9, no. 3, pp. 20–23, 1996.
[24] S. D. Reis, Y. Hu, A. Babino, J. S. Andrade Jr, S. Canals, M. Sigman, and H. A. Makse, "Avoiding catastrophic failure in correlated networks of networks," *Nature Physics*, vol. 10, no. 10, p. 762, 2014.
[25] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman, "Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 17, no. 2, p. 026103, 2007.
[26] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, "Networks formed from interdependent networks," *Nature physics*, vol. 8, no. 1, p. 40, 2012.

[27] X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, “Robustness of interdependent networks under targeted attack,” *Physical Review E*, vol. 83, no. 6, p. 065101, 2011.
[28] C. M. Schneider, N. Yazdani, N. A. Araújo, S. Havlin, and H. J. Herrmann, “Towards designing robust coupled networks,” *Scientific reports*, vol. 3, p. 1969, 2013.
[29] J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, “Robustness of a network formed by n interdependent networks with a one-to-one correspondence of dependent nodes,” *Physical Review E*, vol. 85, no. 6, p. 066134, 2012.
[30] G. Dong, J. Gao, R. Du, L. Tian, H. E. Stanley, and S. Havlin, “Robustness of network of networks under targeted attack,” *Physical Review E*, vol. 87, no. 5, p. 052804, 2013.
[31] B. Bollobás, “Random graphs,” in *Modern graph theory.* Springer, 1998, pp. 215–252.
[32] D. B. West *et al.*, *Introduction to graph theory.* Prentice hall Upper Saddle River, 2001, vol. 2.
[33] B. Bollobás, *Modern graph theory.* Springer Science & Business Media, 2013, vol. 184.
[34] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, “Breakdown of the internet under intentional attack,” *Physical review letters*, vol. 86, no. 16, p. 3682, 2001.
[35] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.
[36] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Social networks*, vol. 1, no. 3, pp. 215–239, 1978.
[37] N. E. Friedkin, “Theoretical foundations for centrality measures,” *American journal of Sociology*, vol. 96, no. 6, pp. 1478–1504, 1991.
[38] G. Sabidussi, “The centrality index of a graph,” *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.
[39] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, “Identification of influential spreaders in complex networks,” *Nature physics*, vol. 6, no. 11, p. 888, 2010.
[40] L. C. Freeman, “Some antecedents of social network analysis,” *Connections*, vol. 19, no. 1, pp. 39–42, 1996.
[41] B. Wellman, “The school childâĂŹs choice of companions,” *The Journal of Educational Research*, vol. 14, no. 2, pp. 126–132, 1926.
[42] P. Erds and A. Rényi, “On the evolution of random graphs,” *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, pp. 17–61, 1960.
[43] R. M. May, *Stability and complexity in model ecosystems.* Princeton university press, 2001, vol. 6.
[44] S. A. Kauffman, “Metabolic stability and epigenesis in randomly constructed genetic nets,” *Journal of theoretical biology*, vol. 22, no. 3, pp. 437–467, 1969.
[45] R. Albert, H. Jeong, and A.-L. Barabási, “Error and attack tolerance of complex networks,” *nature*, vol. 406, no. 6794, p. 378, 2000.
[46] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, “The large-scale organization of metabolic networks,” *Nature*, vol. 407, no. 6804, p. 651, 2000.
[47] A. Wagner and D. A. Fell, “The small world inside large metabolic networks,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 268, no. 1478, pp. 1803–1810, 2001.
[48] M. E. Newman, “The structure and function of complex networks,” *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.
[49] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2003, pp. 137–146.
[50] K. Marzullo, “Tolerating failures of continuous-valued sensors,” *ACM Transactions on Computer Systems (TOCS)*, vol. 8, no. 4, pp. 284–304, 1990.
[51] S. Sahni and X. Xu, “Algorithms for wireless sensor networks,” *International journal of distributed sensor networks*, vol. 1, no. 1, pp. 35–56, 2005.
[52] B. Ao, Y. Wang, L. Yu, R. R. Brooks, and S. Iyengar, “On precision bound of distributed fault-tolerant sensor fusion algorithms,” *ACM Computing Surveys (CSUR)*, vol. 49, no. 1, p. 5, 2016.
[53] H.-H. Wu and M.-Y. Yeh, “Influential nodes in a one-wave diffusion model for location-based social networks,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining.* Springer, 2013, pp. 61–72.
[54] C. Zhang, L. Shou, K. Chen, G. Chen, and Y. Bei, “Evaluating geo-social influence in location-based social networks,” in *Proceedings of the 21st ACM international conference on Information and knowledge management.* ACM, 2012, pp. 1442–1451.
[55] W.-Y. Zhu, W.-C. Peng, L.-J. Chen, K. Zheng, and X. Zhou, “Modeling user mobility for location promotion in location-based social networks,” in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 2015, pp. 1573–1582.
[56] P. Domingos and M. Richardson, “Mining the network value of customers,” in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2001, pp. 57–66.
[57] R. R. Brooks and S. S. Iyengar, “Robust distributed computing and sensing algorithm,” *Computer*, vol. 29, no. 6, pp. 53–60, 1996.
[58] P. Bonacich, “Factoring and weighting approaches to status scores and clique identification,” *Journal of mathematical sociology*, vol. 2, no. 1, pp. 113–120, 1972.
[59] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, pp. 35–41, 1977.
[60] L. Lü, Y.-C. Zhang, C. H. Yeung, and T. Zhou, “Leaders in social networks, the delicious case,” *PloS one*, vol. 6, no. 6, p. e21202, 2011.
[61] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[62] F. Radicchi, S. Fortunato, B. Markines, and A. Vespignani, "Diffusion of scientific credits and the ranking of scientists," *Physical Review E*, vol. 80, no. 5, p. 056103, 2009.

[63] S. H. Lee, P.-J. Kim, Y.-Y. Ahn, and H. Jeong, "Googling social interactions: Web search engine based social network construction," *PLoS One*, vol. 5, no. 7, p. e11233, 2010.

[64] A. Zeng and C.-J. Zhang, "Ranking spreaders by decomposing complex networks," *Physics Letters A*, vol. 377, no. 14, pp. 1031–1035, 2013.

[65] N. Barbieri, F. Bonchi, and G. Manco, "Topic-aware social influence propagation models," in *Data Mining (ICDM), 2012 IEEE 12th International Conference on.* IEEE, 2012, pp. 81–90.

[66] W.-Y. Zhu, W.-C. Peng, L.-J. Chen, K. Zheng, and X. Zhou, "Exploiting viral marketing for location promotion in location-based social networks," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, no. 2, p. 25, 2016.

[67] Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2010, pp. 1039–1048.

[68] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks," *Physica a: Statistical mechanics and its applications*, vol. 391, no. 4, pp. 1777–1787, 2012.

[69] X. Zhang, J. Zhu, Q. Wang, and H. Zhao, "Identifying influential nodes in complex networks with community structure," *Knowledge-Based Systems*, vol. 42, pp. 74–84, 2013.

[70] G. Li, S. Chen, J. Feng, K.-l. Tan, and W.-s. Li, "Efficient location-aware influence maximization," in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data.* ACM, 2014, pp. 87–98.

[71] F. D. Malliaros, M.-E. G. Rossi, and M. Vazirgiannis, "Locating influential nodes in complex networks," *Scientific reports*, vol. 6, p. 19307, 2016.

[72] Z. Wang, C. Du, J. Fan, and Y. Xing, "Ranking influential nodes in social networks based on node position and neighborhood," *Neurocomputing*, vol. 260, pp. 466–477, 2017.

[73] Z. Wang, Y. Zhao, J. Xi, and C. Du, "Fast ranking influential nodes in complex networks using a k-shell iteration factor," *Physica A: Statistical Mechanics and its Applications*, vol. 461, pp. 171–181, 2016.

[74] R. Baral, D. Wang, T. Li, and S.-C. Chen, "Geotecs: exploiting geographical, temporal, categorical and social aspects for personalized poi recommendation," in *Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on.* IEEE, 2016, pp. 94–101.

[75] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management.* ACM, 2014, pp. 739–748.