Chapter 14
Maximizing Social Influence in Real-World Networks—The State of the Art and Current Challenges

Radosław Michalski and Przemysław Kazienko

Abstract The following chapter aims to present the current research in the area of modelling and maximizing social influence in networks. Apart from describing the most popular models for this process, it focuses on presenting the advances in maximizing the spread of influence in social networks. Since most of the research was suited for static networks case, nowadays it is necessary to move it toward the networks that are everywhere around us—the dynamic ones. As is widely agreed in the scientific community, static networks are unacceptable simplification of the real world processes, so current research is moving toward the temporal networks. It is especially important when modelling propagation phenomena, such as the spread of influence, epidemics or diffusion of innovations. In this chapter it is presented how the research on maximizing the spread of influence is starting to explore real-world cases and how the early attempts of solving this problem for temporal networks look like. Moreover, it is shown how to benefit from the temporal properties of the social network in order to achieve better results for spread of influence compared to the static approach.

14.1 Introduction

Social networks are built by humans. And despite the fact that we are predictable to some extent, the social networks we build are in fact dynamic. The factors behind the dynamics may be of different nature, such as meeting new people, changing attitude towards others, switching the job, moving from one place to another and so on. Moreover, the intensity of contacts is also varying in time. In fact, the most accurate method of representing humans communication is the precise information about who contacted whom at which time. By having that it is possible to trace

R. Michalski (✉) · P. Kazienko
Department of Computational Intelligence, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: radoslaw.michalski@pwr.edu.pl

P. Kazienko
e-mail: kazienko@pwr.edu.pl

D. Król et al. (eds.), *Propagation Phenomena in Real World Networks*,
Intelligent Systems Reference Library 85, DOI 10.1007/978-3-319-15916-4_14

the information or influence flow in social networks. Yet, without knowing about the content or the essence of the communication, only some assumptions about potential paths of information diffusion may be made. Moreover, as a single contact entry most often contains just two entities: sender and recipient (directed case) or contacting parties (undirected case), by using this level of granularity it is impossible to benefit from the mature apparatus of social network analysis (SNA), since no graph exists yet. It also requires a lot of storage comparing to another method—building a time-aggregated social network from the contact logs. In this case the information about communication gets aggregated and the intensity of contacts is most typically expressed as weights over edges (contacts) between nodes (individuals), as presented in [6]. This approach allows to obtain a broader view of interconnections in networks, distinguish groups, hubs, nodes on boundaries of the network and lets to perform other analyses that are offered by SNA techniques [13]. Unfortunately, while the time-aggregated view of the network is used, from an information or influence propagation point of view [43, 44], the most important aspect is missing: the order of contacts that is crucial in analysing the flow of information. As it was stated in [62], in these networks one assumes transitive paths, and this assumption does not hold in temporal or most granular representations of social networks. Moreover, as the contacts within social networks are often *bursty* [4], the static representation of networks will also ignore this fact leading to wrong conclusions about the dynamic processes taking place there. This is especially crucial when modelling the spread of epidemics, since the accuracy of predictions may strongly influence the potential actions in healthcare [53]. To not to loose the temporal information, researchers more and more often use temporal representation of networks and a comprehensive overview of methods of building temporal networks may be found in [34]. In this work the authors state that the literature on static graphs is many times larger than on temporal graphs and this is for a natural reason: it is much easier to analyse static graphs, especially analytically. Naturally, it is not a reason to avoid this direction, since as the research reveals, only temporal networks are representing the surrounding world accurately—the static approach is considered as leading to wrong conclusions about dynamical processes.

Looking from the perspective of social influence, this process has enough psychological and sociological complexity itself and because of that it shouldn't be analysed by using simplified underlying layer. It is modelled upon sociological assumptions about how people become influenced [32] and using time-aggregated networks to model humans' interactions definitely does not help in understanding the speed and directions of spread of influence. As it will be presented later in this chapter, most of traditional methods for analysing the influence processes in social networks base on a static representation, especially in the area of maximizing the spread of influence [39]. Since in this work the problem of influence maximization has been proven to be NP-hard, many heuristics were proposed, but mostly for time-aggregated networks. Nowadays the direction of research on modelling dynamic processes in networks should consider dynamic networks as a base, since by simplifying the reality the obtained results and drawn conclusions may be wrong.

The goal of this work is to present the state of the art in the area of maximizing the spread of influence in social networks, the limitations of it in the static networks and recent trends in using the dynamics of networks or past propagations to obtain better results. To achieve this, most common models of the social influence are presented in Sect. 14.2. Section 14.3 shows what are the variants of the challenges in social influence in networks, since the influence maximization is not the only one. Next, methods for maximizing the influence in static networks are introduced in Sect. 14.4. Then the concept of *Temporal Social Networks* is introduced and most typical representations of them are presented in Sect. 14.5. Having these introduced, Sect. 14.6 reveals the experimental study results showing that the temporal approach may outperform the static one for simple heuristics when considering the temporal underlying layer for the spread of influence. Section 14.7 comes back to presenting the state of the art, but presents how researchers try to take the advantage from the network dynamics or history to obtain better results in the challenge of maximizing the spread of influence.

14.2 Modelling the Spread of Influence

Before presenting the most popular models of social influence, it is worth to quote one sentence from [32] by Watts and Dodds:

> (...) it is still the case that formal models of social influence suffer from a dearth of realistic psychological assumptions.

The problem of fitting the real-world data to models and trying to answer the question whether particular influence processes may be modelled with a chosen approach is still challenging. It lies in the complexity of human behaviour and the impossibility of separating social processes that are occurring simultaneously. Still, many results achieved in this area tend to contradict this pessimistic point of view of Watts and Dodds and continuous development of models or models' variations suggests that models will fit the reality even better in next few years [2]. On the other hand, there still remains the gap between formal models and psychological explanation that requires to be intensively studied to find the psychological rationale of particular behaviour expressed in these models.

Since the strength of social influence depends on many factors such as the intensity of relationships between people in the networks, the network distance between users, temporal effects, characteristics of networks and individuals in the network [69], it is relatively hard to model all these factors combined. However, vast of research shows that under some assumptions there exist models that fit the reality well. Below the most important models that are most commonly used in this area are presented: *the Linear Threshold model*, *the Independent Cascade model*, *the Voter Model*, and *the Naming Game*. Each of them incorporates the sociological background of the influence process, but as was previously stated, sometimes it is just a loose interpretation of humans' behaviour, that, luckily, still fits the reality well for some cases. For these models their recent variants which are suitable for real-world scenarios are described as well.

From the historical perspective, studying the social influence in terms of analytical process was the case of trying to model how the influence spreads in time. Starting from a set of influenced nodes in time t_0 which are in this work denoted as $\Phi(0)$, as time unfolds, more and more of neighbours of $\Phi(0)$ become influenced if they fulfil the model criteria. Most typically, these processes are modelled in directed graphs and focus on a *progressive* case, where nodes may become *influenced* from *uninfluenced* state, not the other way round [39]. Since this is a network approach, the influence process occurs through edges in graph and most typically no other external factors of influence are considered, such as out-of-network sources.

14.2.1 The Linear Threshold Model

The most recognizable model for social influence is Granovetter's Linear Threshold model [31]—*LT*, but similar approach was also proposed in [66]. In this model, a node v is under influence of its influenced neighbours w denoted as N_v^{inf} according to a weight $b_{v,w}$, such that $\sum_{w \in N_v^{inf}} b_{v,w} \leq 1$. Each node v has a *threshold* θ_v from the interval [0, 1] and this threshold represents the level that has to be met by the aggregated sum of v's neighbours influence weights in order to influence the node v. So the formal condition of influencing the node v is as follows:

$$\sum_{w \in N_v^{inf}} b_{v,w} \geq \theta_v. \tag{14.1}$$

The influence process ends where more nodes cannot be influenced—this is the formal stop condition for the static case. The way how this model works is presented in Fig. 14.1, while formally it is presented as Algorithm 14.1 for the case of uniformly assigned threshold values θ_v (based on [76]).

Here, the value of θ_v represents the individual's chances of becoming influenced when its neighbours are influenced. So all the psychological factors are included in this parameter and it should be also underlined that this approach represents the individual's perspective rather than the influencer perspective. Granovetter illustrated the model with the hypothetical case of a riot. Since individuals were unsure what are the costs and benefits of joining it, they observed their peers and considered joining only when sufficiently many of their neighbours joined the riot, otherwise they refrained.

Of course, the biggest question is how to assign particular values of θ to individual nodes and there are two most typical approaches: draw them from a probability distribution $f(\theta)$ as introduced in [31] or hard-wiring them at a fixed value [7, 61]. The most interesting and realistic scenario is the former one, i.e. drawing θ_v from a distribution, since the distribution represents both the average tendencies and also the heterogeneity present in the population. Lowering or raising the mean

Algorithm 14.1 Linear Threshold model

Require: Graph $G(V, E)$, set of initially influenced nodes $\Phi(t_0)$
return Final set of influenced nodes $\Phi(K)$
k = 0;
Uniformly assign random thresholds θ_v from the interval [0, 1];
while $k = 0$ or $\Phi(t_{k-1}) \neq \Phi(t_k)$ **do**
  $\Phi(t_{k+1}) = \Phi(t_k)$;
  $uninfluenced = V \setminus \Phi(t_k)$;
  for all $v \in uninfluenced$ **do**
    if $\sum_{w \text{ influenced neighbour of } v} b_{v,w} \geq \theta_v$ **then**
      influence v;
      $\Phi(t_{k+1}) = \Phi(t_{k+1}) \cup \{v\}$;
    end if
  end for
  $k = k + 1$;
end while
$\Phi(K) = \Phi(k)$;
Return $\Phi(K)$;

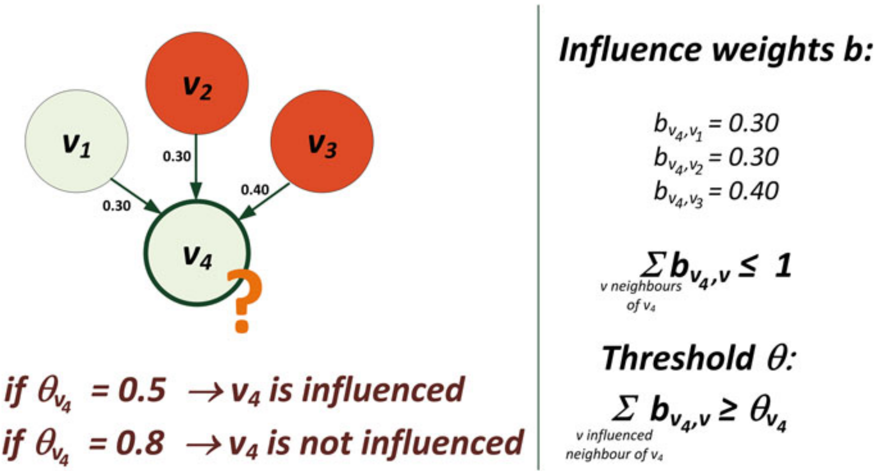


Fig. 14.1 The illustration showing how the LT model works

of $f(\theta)$ would modify the general susceptibility of the population, while increasing or decreasing the variance would correspond to an increase or decrease in variability in susceptibility across individuals [32]. Still, hard-wired thresholds are also often considered in the research. An exemplary spread of influence process following the *LT* model is presented in Fig. 14.2.

It should be underlined that this process is time-independent, since it considers iterations rather than time. However, in most research works in this area an iteration represents a single time step, this is why the notation of $t_0, \ldots, t_K$ is often used instead of iterations $i_0, \ldots, i_K$.

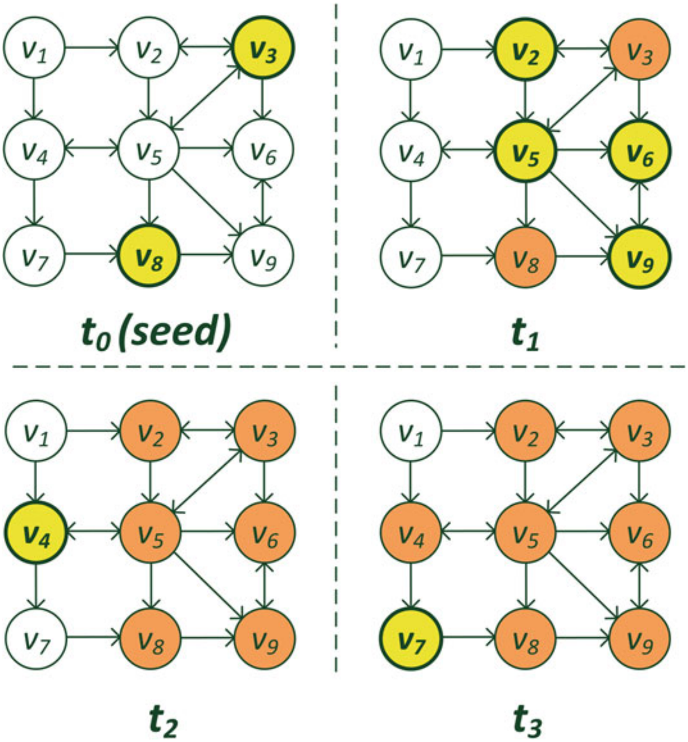


Fig. 14.2 An exemplary social influence process following the linear threshold model in graph G. The threshold value is fixed for all nodes, $\theta_v = 0.33$. At the beginning $\Phi(0) = \{v_3, v_8\}$, at the end of the process $\Phi(t_3) = V(G) \setminus \{v_1\}$, since v_1 cannot be influenced for this model parameters. Nodes in *bold* were influenced at this particular process step

The following theorem and proof are excerpted from [39] and show the NP-hardness of influence maximization problem for the LT model. In the same work the proof for NP-hardness of the same problem for IC model is shown.

Theorem 14.1 *The influence maximization problem is NP-hard for the Linear Threshold model.*

Proof Consider an instance of the NP-complete *Vertex Cover* problem defined by an undirected n-node graph $G = (V, E)$ and an integer k; it is expected to find a set S of k nodes in G so that every edge has at least one endpoint in S. It is shown that this can be viewed as a special case of the influence maximization problem. Given an instance of the *Vertex Cover* problem involving a graph G, a corresponding instance of the influence maximization problem by directing all edges of G in both directions is defined. If there is a vertex cover S of size k in G, then one can deterministically make $\sigma(A) = n$ by targeting the nodes in the set $A = S$; conversely, this is the only way to get a set A with $\sigma(A) = n$. $\square$

The LT model became a core of many modifications or extensions. For instance, [26] extends this model by introducing temporal decay, as well as factors such as the influence-ability of a specific user, and influence-proneness of a certain action. On the other hand, [5] proposes topic-aware extensions of *LT* model. In [60] the authors consider multiple cascades of *LT* model and they allow nodes to switch between them, whereas [10] introduces a number of modifications to the competing model variant: the authors force nodes to draw one cascade they join at the end of the process or consider the mutual influence of cascades on each other.

14.2.2 The Independent Cascade Model

The next model has its roots in interacting particle systems [23, 49] and is called the Independent Cascade model—*IC* [25, 39]. Again, the process starts with a set of influenced nodes $\Phi(0)$, but each node v in the network has a probability $p_{v,w}$ assigned. According to this probability, the node v gets a single chance to influence its neighbour w and when it fails, it will have no other chance. If it succeeds, w will become influenced in the next time step. Similarly to the LT model, the process runs until no more influences are possible.

From psychological perspective, in this model the influencer becomes more important, since he or she holds the probability $p_{v,w}$. This is one of the major differences between *IC* and *LT* models. In the *LT* model, the influence process parameter was assigned to the uninfluenced node and in the *IC* model it is hold by the potential influencer. Just as in the previous model, the probability may be fixed or drawn from a distribution $f(p)$. The Independent Cascade model is presented in Algorithm 14.2 (based on [76]).

Algorithm 14.2 Independent Cascade model

Require: Graph $G(V, E)$, set of initially influenced nodes $\Phi(t_0)$, activation probabilities $p_{v,w}$
return Final set of influenced nodes $\Phi(K)$
k = 0;
while $\Phi(t_k) \neq \{\}$ **do**
  $k = k + 1$;
  $\Phi(t_k) = \{\}$
  for all $v \in \Phi(t_{k-1})$ **do**
    for all w neighbour of v, $w \notin \cup_{j=0}^{k} \Phi(t_j)$ **do**
      $rand$ = generate a random number in [0, 1];
      if $rand < p_{v,w}$ **then**
        influence w;
        $\Phi(t_k) = \Phi(t_k) \cup \{w\}$;
      end if
    end for
  end for
end while
$\Phi(K) = \cup_{j=0}^{k} \Phi(t_j)$;
Return $\Phi(K)$;

Again, there are many variants of the *IC* model. Already mentioned work of [5] introduces the topic-aware approach also for this model, while [40] study the *decreasing cascade model*. One of the problems with the base *LT* and *IC* models is that they do not provide the influence probabilities and there are works that try to obtain them from past propagations. It may not be considered as a extension of a base model, but a way to make the probabilities or threshold more realistic. One of the works in this area is [65], but this topic will be covered in Sect. 14.7.1. There also exists the approach to model multiple independent cascades in the network [8].

14.2.3 The Voter Model and the Naming Game

An interesting case of influence in networks is the situation where two separate opinions or influences are competing in the society. This phenomenon may be observed in many situations and it has its roots in studying the consensus processes [51] or the language dynamics [20]. Below there are two variants of the process presented: the Voter Model (*VM*) and the Naming Game (*NG*).

The Voter Model introduced in [19] and extensively analysed later in [33] assumes that each node in the network can hold one of two opinions and by interacting with others it may switch the opinion to the opinion of the peer. This model introduces also the degree of conformity which defines whether a node will follow the majority (conformist) or minority (non-conformist), see [37].

On the other hand, the Naming Game, also referred to as binary-agreement model [74], introduces another variant of forming the opinion or spreading the influence. At any time a node may possess one of two competing opinions or two opinions simultaneously. In a given time step, we choose a node randomly, designate it as a speaker and choose one of its neighbours randomly and it is a listener. The speaker proceeds to convey its opinion to the listener (chosen randomly if it possesses two) to the listener. If the listener possesses this opinion already, both speaker and listener retain it while eliminating all other opinions; otherwise, the listener adds the opinion to his list [73].

Both of these models are useful in studying common phenomena occurring in social networks that involve binary options, such as reaching the consensus on contradictory opinions or observing which of competing parties will win the election. The current research trends suggest that these models will be actively studied and extended in the future [48, 52, 58, 64, 77].

14.2.4 Summary

The above presented models are just a selection of models that allow to analytically study the influence processes in social networks. As it was presented, they differ by the perspective (*LT* vs. *IC*), by the number of competing influences (*VS* and *NG* vs. others), but all of them are linked to the same process—spread of influence.

Sometimes their applicability is limited, but as far the empirical research shows, they model the human behaviour accurately in some cases, even if the psychological background of an individual is more complex than just a single parameter.

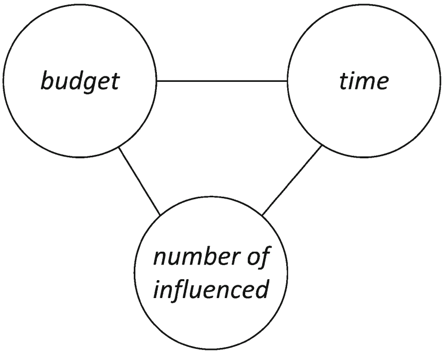
14.3 Social Influence Challenges

The work of Goyal et al. [28] brings some insights into the problem of influence in social networks. Since most often only the case of maximizing the spread of influence with a given budget k was considered, there are some other research questions in this area. One should ask about minimizing the time of influence (number of iterations to influence a given number of nodes by having the budget k) or about minimizing the budget k to influence a given set of nodes. The generalization of the challenges in this area may be considered as a constrained optimization problem, as presented in Fig. 14.3. The dimensions that can be optimized are the budget, the time and the number of influenced. Here, only one or two dimensions may be constrained and the third is optimized. Below, each of the dimensions is briefly described to show how they are understood by most of the researchers.

14.3.1 Budget

The *budget* in the spread of influence problem in social networks is considered as an amount of the resource that can be spent on influencing nodes (please note that some literature prefers the more general term *activating*, e.g. see [39]). This resource is most often expressed as a budget k of different nature such as money, gifts, conversations. However, each successful influence of a node in the network reduces the budget. Typically, it is assumed that the amount of a budget taken for influencing a node is equal for all of the nodes in network. Sometimes it may be

Fig. 14.3 The optimization problem for the influence in social networks



true, e.g. if in a marketing campaign the same product is being sent to different customers the cost of distributing the product among them is considered to be equal. On the other hand, as the influence process is a subjective one, even by spending some amount of the budget on a user, he or she may not become influenced and this susceptibility may differ from user to user. As it was already presented, those psychological aspects are included in the spread of influence model properties (such as θ for the LT model or $p_{v,w}$ for the IC model), but these models do not consider the varying cost of influencing individuals in the set of seeds $\Phi(0)$. However, for the problem stated in this chapter, as well as in the research in this area, it is assumed that the cost of initial influencing a node is equal for all the nodes.

14.3.2 Time

The time constraint expressed here means that we want to influence nodes in a given time and to evaluate the results of different methods this particular time is considered as a stop condition. Typically, the models work until no more nodes could be influenced. This is a natural stop condition and it is reasonable for static networks and described models. Of course, when the time constraint appears, the process may be evaluated sooner. But when considering the temporal social networks, the use of this hard stop condition may become more complicated, since if the network changes, the influence process may be infinite, e.g. if with every iteration new nodes join the network it could be hard to say at which moment the algorithms should stop. This is why the time constraint may be crucial for temporal social networks, since it introduces a moment which allows to compare methods.

From the marketers perspective, if they spent some budget on influencing nodes, they want to get the return from this investment in a given time, e.g. they want to have people interested in buying the product when it is offered and not discontinued [16]. On the other hand, other examples may be not so focused on the time dimension. For instance, spreading good manners among the society is one of the examples. Naturally, the sooner the habits will improve the better, but the time aspect is not so important here compared to the whole success of the campaign. This is why the use of time constraint is sometimes desired, but sometimes this dimension is being left unconstrained. However, reader should have in mind that for temporal social networks the use of time constraint is somehow natural, since the stop condition that no new nodes will become influenced may be wrong.

14.3.3 Number of Influenced

The last, but most often considered dimension is the *number of influenced*. The number of influenced means how many nodes were influenced or activated in the process. From the historical perspective, this dimension was the one that was maximized while the other two left constrained or unconstrained, but [28] started to consider some other variations of the problem, e.g. by trying to minimize the time of influencing a given number of nodes.

However, when analysing advancements in research, the problem of maximization the influence is the most popular one among researchers.

14.4 Maximizing the Spread of Influence

The problem of finding most influential seeds in a social network was originally stated in [22]. The authors posed a question on how to pick nodes and *influence* them with some idea to maximize the overall spread of this idea across network. In this work the example of a marketing campaign was used and the researchers considered the network value of a customer, i.e. the benefits for the company if this customer will influence its neighbours. In fact, nowadays selling techniques often base on viral marketing, so it seems that the business strongly believes in a potential of such an approach. The influence of nodes on each other was modelled as a Markov random field [41] and obtained results revealed that this approach may be promising. In the next work on this topic the authors used a linear model where the solution for influence maximization based on solving linear equations [63]. However, what this model lacked for, was the iterativeness, since it reflected the joint distribution over all nodes. Compared with the psychological research on social influence or diffusion of innovation, the process is rather iterative, so the models representing it are most often of this kind.

14.4.1 The Greedy Algorithm

The work that showed different approach to the one presented by Domingos and Richardson was [39]. The authors started by assuming that the influence is more an iterative process, so they analysed two models of this kind, namely LT and IC. By basing on these, firstly they considered the hardness of the influence maximization problem and in both cases it was proved to be NP-hard, see Theorem 14.1 with the corresponding proof for LT model and [39] for the proof for IC model. Then, by taking the advantage of the properties of submodularity [67] and the research on greedy hill-climbing algorithm they show that the greedy method may outperform classic approaches based on network measures, such as top degree or top betweenness. In fact the authors show that the outcomes of this approach may be not worst than 63 % of the optimal solution. The greedy method pseudo-code is presented as Algorithm 14.3 (based on [18]). Here, given a social network $G = (V, E)$ consisting of sets of vertices and edges, an initial seed set of k nodes is being chosen iteratively which maximizes the influence. In each step of the algorithm a single vertex is chosen, such that the influence of the set Φ and this vertex is the greatest. Unfortunately, this algorithm has few drawbacks. One is the efficiency, since the influence is estimated with R simulation steps [18]. The other is that the algorithm is trying to pick nodes that maximize the influence in each iteration. When comparing it to the chess game, it always chooses the move that is giving the best position at the moment, not thinking about the next move. Sometimes it may be better to look at a combination of moves or nodes rather than a single next move to maximize the overall result and the greedy algorithm avoids it by its nature, since it finds the local optimum. However it still provides acceptable results comparing to the optimal solution, but sacrificing the efficiency.

Algorithm 14.3 Greedy algorithm for maximizing the influence

initialize $\Phi(0) = \emptyset$ and $R = 10000$
for $i = 1$ to k **do**
  for each vertex $v \in V \setminus \Phi(0)$ **do**
    $s_v = 0$
    for $i = 1$ to R **do**
      $s_v = |Inf(\Phi(0) \cup \{v\})|$
    end for
    $s_v = s_v / R$
  end for
  $\Phi(0) = \Phi(0) \cup \{\arg\max_{v \in V \setminus \Phi(0)} \{s_v\}\}$
end for
output S

To overcome the drawbacks mentioned above, the research splits in two directions. Firstly, a number of techniques were proposed to optimize the greedy algorithm. Secondly, researchers started to search for new ways of maximizing the spread of influence. Below it is presented how the greedy algorithm was improved and later on new ideas on maximizing the influence in social networks are introduced.

14.4.2 Greedy Algorithm Optimization

In the work of Leskovec et al. [46] there is also one more drawback of the greedy algorithm shown. As for now it was assumed that the cost of acquiring a single node is equal to others, but in social networks it may not be the case. For instance, for an ongoing marketing campaign its designers like to give to influential social network users some incentives to end up with higher spread of influence, their expectations for value of incentive may vary from one to another. On the other hand, there are some scenarios where the same products are being sent to different users assuming that they become influenced, so the potential cost is equal. However, since the influence process is a subjective rather than an objective one, the same gift may not result with the same satisfaction of users from the product. In contrast, there are some cases of social networks where the equal cost is possible. For instance, in epidemiology it is often assumed that cost or probability of infecting a person is the same for all the population, but this is not the case of influence. In sensor or computer networks also dealing with devices may introduce the same cost for each of them, but the assumption of the same cost for different users seems to be limited.

The mentioned work [46] analysed the case of different cost for influencing each node and proved that with this assumption the greedy algorithm performs badly. To overcome this limitation the authors introduce a novel approach, *Cost-Effective Forward selection* (*CEF*) that uses the greedy algorithm and the cost-sensitive method in parallel and the results of these methods are compared later to find the better one that will be used. Moreover, again by using the submodular properties of the cost function, the researchers are able to reduce the number of possible runs of evaluation of the quality of selected node ($Inf(\Phi \cup \{v\})$), because they base on the fact that the marginal increase of benefits with each added node does not increase more than in previous evaluation. This approach is called by them *Cost-Effective Lazy Forward selection* (*CELF*) and comparing to the greedy algorithm is up to 700 times faster with still acceptable results of at least $\frac{1}{2}(1-\frac{1}{e})$ of optimal solution. Comparing it with the greedy algorithm, which proposes $(1 - \frac{1}{e} - \varepsilon)$, makes CELF a very good rival.

However, at least for the IC model, there was still space for improvement, as shown in [18]. Here, the authors decided to base on random graphs in order to reduce the number of runs R (see Algorithm 14.3). Their approach assumes that for IC model it is possible to reduce the graph of influences to only these edges that are potentially reachable from the set Φ at ith iteration. This change allows to gain additional 15–34 % improvement in running efficiency by keeping the same level of quality. More interestingly, in the same work the researchers propose new heuristic that significantly improves the influence spread while running more than six orders of magnitude faster than all greedy algorithms—*DegreeDiscountIC*. This approach is basing on a degree heuristic, but *discounts* the degree of a considered as a seed node v by the value of already influenced neighbours of v, since there exists a non-zero probability that this node will become influenced by one of its influenced neighbours and it makes it less attractive as a seed.

Goyal et al. shown that further exploitation of submodularity may lead to even better results for greedy algorithms. In [29] there is an extension of CELF algorithm that leads to at least 35 % gain in performance. The idea is to store a heap for all non-selected nodes that contains the information not only about a marginal gain of particular node, but also the marginal gain of the best node from these evaluated before this node. Due to this trick there is no need to recalculate marginal gain of a node if this node was not selected resulting in less iterations of an algorithm.

14.4.3 Avoiding Greedy Search

One of the approaches was already mentioned, it was the *DegreeDiscountIC* algorithm [18]. The same authors proposed another method [17], *Maximum Influence*

Arborescence (MIA), which also exploits the submodularity. However, in this case for the IC model, for each pair of nodes maximum influence paths are calculated. Then paths below a specified threshold of influence are discarded, focusing only on local regions of influence. Afterwards, these paths creating tree structures which do need to be updated often are joined and the calculation of the influence spread may be done recursively. This leads to significant gains in effectiveness of the algorithms comparing to others and introduces no loss in terms of quality. Moreover, the threshold parameter may be interpreted as a way of controlling the time of influence and the overall spread.

Another work [38] also avoids the greedy approach while outperforming it in terms of speed (2 to 3 orders of magnitude) and bringing similar results in accuracy. Authors claim that this is the first approach that uses simulated annealing [55] in solving the problem. In the case of influence maximization this approach starts with random seeds and then tries to move in the space of possible solutions (initial seeds) towards the local minimum by swapping at most one node in the seed set until the stop condition will be applied.

An interesting insight into the problem was given by Shakarian and Paulo in [68] where the authors propose an algorithm that guarantees to activate (influence) the whole network. The solution does not find the minimal set of seeds, but its outcomes may be compared to the budget k—if the seed set is less or equal k, the algorithm will fulfil the requirements and, moreover, the whole network will be influenced. The approach base on removing edges in the graph (by basing on the idea of shell decomposition [12]), but it also guarantees to influence the whole network.

The last mentioned algorithm in this section is *Simpath* that is intended to maximize the influence for the LT model [30]. This algorithm is operating on paths of influence in the social network by assuming that most of the influence is local. Results reveal that the algorithm outperformed the MIA method, considered as the state of the art in the task of influence maximization for static networks [17].

14.4.4 Summary

All the above presented techniques may be considered as purely structural ones. In here, researchers do not use any kind of attributes of nodes other than their structural properties, such as location in the network or interconnectivity with other influenced nodes. The only parameter that may differentiate the nodes is the cost of influence, but in most cases it was assumed to be uniformly distributed. This approach of basing just on network structural properties makes the proposed algorithms universal, since they do not base on any network-specific attributes. As it will be shown later, there is also another emerging direction in this research that is basing on the data, so the merit of the social network communication. Moreover, it is worth emphasizing how many of the presented approaches took the advantage of the submodularity property.

However, all algorithms presented in this section suffer from one drawback which makes them just a rough simplifications of reality. Here, the network dynamics is not considered, and as it was already stated in Sect. 14.1, the ignorance of this fact may lead to wrong conclusions about the outcomes of the process. Before taking the reader in the area of research which tries to benefit from the dynamics of the network to maximize the influence, in the next section theoretical framework for representing dynamic networks is presented. Then a small empirical study in Sect. 14.6 shows that it is worth to make use from network dynamics, and following this direction current advancements in the topic of maximizing the spread of influence in the dynamic configuration are presented.

14.5 Temporal Social Networks

As it was already mentioned, real world social networks are rarely static. Our interactions occur in an ordered way, sometimes they are bursty, sometimes these contacts are suspended, but nevertheless there is a dynamics embedded. Until the era of IT the problem of gathering the data about interactions was definitely harder than nowadays, so it was another reason (but not the only one) why researchers based mostly on static networks. But when the era of electronic communication begun, it is relatively easy to track human communication, at least for some communication channels, such as e-mails, phone calls, instant messengers or social networking sites interactions. Moreover, there exist projects that try to obtain the data of real-world interactions, e.g. hospital ward contacts [71], conferences' participants [14] or students' behaviour [24]. Now the real research question is how to benefit from this time-annotated data to extend the knowledge on the social influence.

Before trying to find the answer on it, it is worth to discuss how these kind of data should be represented to not to loose the temporal information. One of the most extensive survey work on Temporal Social Networks *TSN* [34] depicts two major approaches in representing the temporal information in social networks, depending on the contacts type. These are enumerated and briefly described below.

14.5.1 Contact Sequences

A *contact sequence* is obtained mostly from communication data where single contact between individuals is timestamped. As a relatively straightforward migration from event logs, it represents actors as nodes and edges between them have the temporal information, i.e. at which time a communication occurred. This approach is especially suitable when the duration of interactions is negligible, so it is preferred when representing asynchronous communication, such as e-mails, text messages and, to some extent, phone calls. The most typical formulation of contact sequences is as follows: for a set of vertices V a contact sequence is a set of triples (v_i, v_j, t) representing contacts between nodes v_i and $v_j \in V$ at time t—see Fig. 14.4.

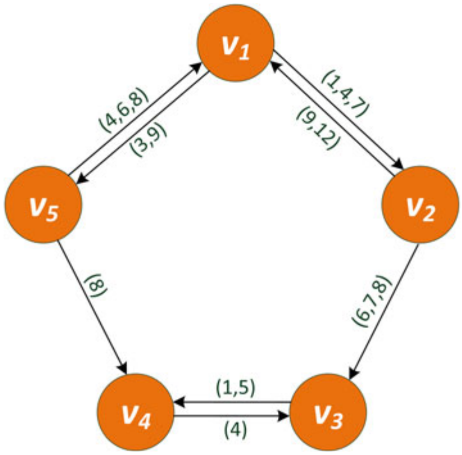


Fig. 14.4 The representation of contact sequences. Here, for instance, node v_1 contacted v_2 in time *1*, *4*, *7* whereas the reverse communication occured in *9*, *12*

Very often this contact sequence is transformed into graphs where contacts in the same timeframe are grouped, making them a sequence of time-ordered static social networks [11], as presented below:

$$\begin{aligned} TSN^m &= \langle T_1, T_2, \ldots, T_m \rangle, && m \in \mathbb{N} \\ T_t &= SN_t(V_t, E_t), && t = 1, 2, \ldots, m \\ E_t &= \langle v_i, v_j \rangle : v_i, v_j \in V_t, && t = 1, 2, \ldots, m. \end{aligned} \tag{14.2}$$

In this formulation *TSN* represents the sequence of static networks SN_t aggregating contacts in timeframe t making it more a evolving static social structure. However, the authors of [34] argue that this representation may miss many important points of temporal activity, what indeed is true, since some simplifications in the contact orders arise. On the other hand this simplification may be helpful in applying most popular models of social influence. The highest level of aggregation which results with a single static social network is often called a *time-aggregated graph*.

14.5.2 Interval Graphs

Another form of temporal networks are *interval graphs*. Here, in opposite to contact sequences, the edge is active in a period of time, rather than it appears at specific time. This kind of temporal networks is more suited for respecting the duration of contacts, so its application is also different than in the former approach. One of the examples might be tracking the duration of interpersonal contacts as the exposure on infection—the longer the exposure, the higher the probability of becoming infected. Here the edges are not active over a set of times but rather over a set of intervals $T_e = \{(t_1, t_1'), \ldots, (t_n, t_n')\}$, where the parentheses mark the periods of activity.

The above representations of temporal networks offer the highest granularity which makes them perfect to track precise interactions between nodes. Naturally, depending on the research goal, some simplifications may be used, but it is worth to remember what consequences particular simplifications introduce. In the next subsection problem of transitivity in temporal networks is presented, since it is important in understanding the limitations of time-aggregated approach in analysing the diffusion or influence processes.

14.5.3 Limitations of the Time-Aggregated Approach

Consider the following graphs: SN_{AGG} which consists of 4 nodes, namely $V_{AGG} = \{v_1, v_2, v_3, v_4\}$. This is a time-aggregated version of obtained contact sequences. Contact sequences were presented as graphs SN_1, SN_2, SN_3, named so, because contacts occurred in times $1, 2, 3$, respectively. For an illustration, see Fig. 14.5, where (a) denotes the time-aggregated graph and (b) contact sequences unfolded to three social networks.

Assuming that in the network a linear threshold process takes place with the threshold set uniformly for all nodes $\theta = 0.5$ and initially only node v_4 is influenced, the process will behave differently for both networks. For the time-aggregated graph after two iterations all nodes will become influenced. For the temporal network in $t = 1$ no new nodes will activate, the same in $t = 2$. In time $t = 3$ the node v_4 will be able to influence its neighbours and still v_2 will be not influenced. This simple example shows that the influence process is prone to the network dynamics and having this in mind the next section presents a simple experimental study showing that it is possible to benefit from it to obtain better influence maximization results.

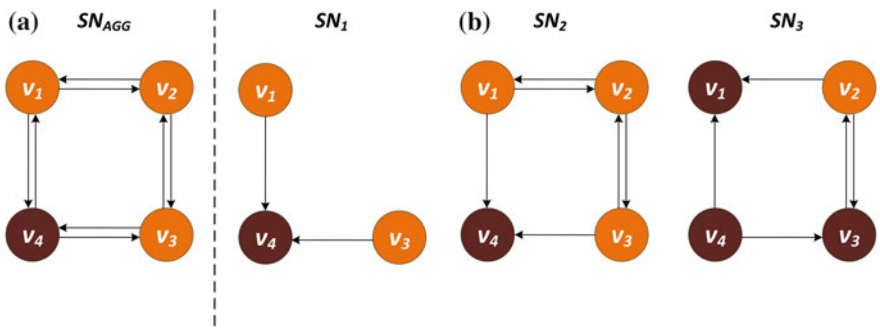


Fig. 14.5 The LT influence process for time-aggregated graph and a temporal network

14.6 Spread of Influence—Temporal Approach

In the research which was presented in [56] it was decided to evaluate how the most typical heuristics based on the network structure perform on temporal and static networks. A special attention has been turned to observation of dynamics of the influence that spreads over temporal network after choosing initial seed sets. The basic heuristics basing on structural network measures were evaluated to see whether time-enhanced versions of these measures will perform better. Since the goal of this chapter is to present the current advancements in the research on maximizing the social influence, the experiments are briefly described here just to show that indeed moving with the spread of influence towards dynamic networks may be beneficial in solving the problem and it places it in real-world setup rather than in an abstract one.

14.6.1 Introduction

The general problem considered in that context is what kind of networks should be used to perform better in seeding and finally in the spread outcome. Two main network kinds have been further studied: static one that aggregates equally all knowledge from the past (TSN^1 in Fig. 14.6) and temporal one that splits the past period into more or less time intervals: TSN^{10} with 10 equal time windows and TSN^5 with 5 time frames. The temporal approach corresponds to dynamic context of seeding, whereas an aggregated social network reflects typical static seeding circumstances. The goal of the experiment was not to focus on seeding strategies itself, but to analyse the process under different assumptions for the aggregation level.

14.6.2 Experimental Setup

14.6.2.1 Time-Dependent Measures

Firstly, three simple aggregations were introduced, which allow to order users based on structural measures (total degree, in-degree, out-degree, betweenness, closeness) respecting all periods in the temporal social network in the accumulated way—maximum, minimum and sum. These aggregations, however, do not make use of sequential nature of time and general phenomena that recent social relationships are likely to be more influential than old ones. Hence, the nine new aggregations that take into account also the “forgetting” aspect of time are introduced. Here, the value of a given structural measure in the most recent time window is the most important, while the measures value in the oldest period is the least valuable. The purpose of this was not only to capture the dynamics of user behaviour but also to emphasize users

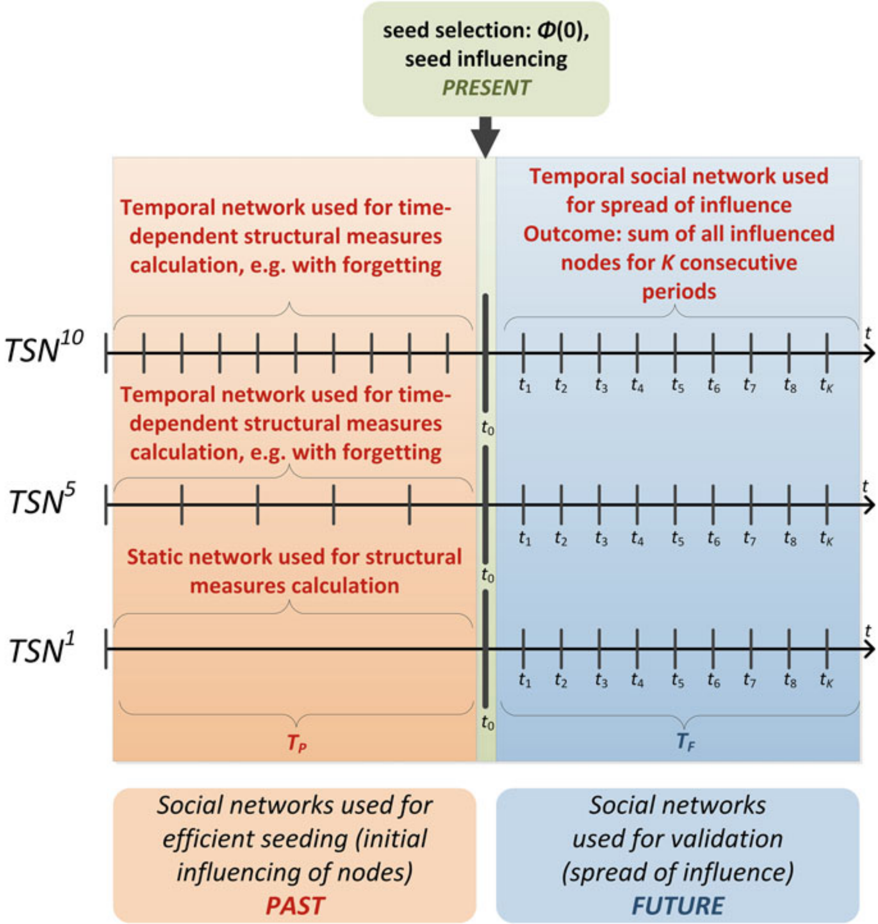


Fig. 14.6 Seeding is performed at present based on the knowledge about past dynamics of the social network (in time T_P). The seed—set $\Phi(0)$ of initially influenced nodes is used to spread of influence in the dynamic social network in the future (in time T_F). Three kinds of 'learning' social networks used in the experiments on seed selection are depicted one below another: TSN^{10} with 10 time windows, TSN^5 with 5 time frames, TSN^1—aggregated-static (one time window)

latest activities. So the new aggregations were applying different kinds of forgetting, e.g. linear, hyperbolic or exponential forgetting.

14.6.2.2 Aggregation Levels and Influence Model

All the aggregations combined with all typical node structural measures (in-degree, out-degree, total degree, betweenness and closeness) where used to create node rankings and select the seed set for spreading the influence. In other words, nodes in the temporal social network from the past were ranked according to the time-aggregated values of their structural measures and this aggregation was performed for all component networks used for seeding, see the left part of Fig. 14.6. Next, 5 % of top ranked nodes were used for seeding, see the middle part of Fig. 14.6. It means that these top nodes form the initial set $\Phi(0)$ of already influenced nodes that may influence others in the following periods, see the right part of Fig. 14.6. In each case, the second part of the dataset was split into ten windows of equal duration, to reflect the dynamic behaviour of the network. As a model of influence, the linear threshold was used and three levels of θ were evaluated—0.33, 0.50, 0.75 assigned uniformly for all nodes.

14.6.2.3 Datasets

The experiments were conducted using five real-world social networks representing the communication between company employees or social services users (Table 14.1). All of them were extracted from communication datasets downloaded from the Koblenz Network Collection (KONECT)[1] repository. Each social network has timestamped edges, so it allowed to perform temporal analysis. The properties of the datasets are presented in Table 14.1.

Table 14.1 Descriptions and basic properties of used datasets

Dataset ID	Network description	No. of nodes	No. of timestamped edges	Period of communication
1	E-mail communication between employees of manufacturing company [57]	167	82,927	2010-01-02 ... 2010-09-30
2	The Enron email network [42]	87,101	1,147,126	1998-11-02 ... 2002-07-12
3	Messages sent between the users of an online community of students from the University of California, Irvine [59]	1,899	59,835	2004-04-15 ... 2004-10-26
4	Facebook user to user wall posts [72]	46,952	876,993	2004-09-14 ... 2009-01-22
5	The reply network of the social news website Digg [21]	30,398	87,627	2008-10-28 ... 2008-11-13

[1] http://konect.uni-koblenz.de.

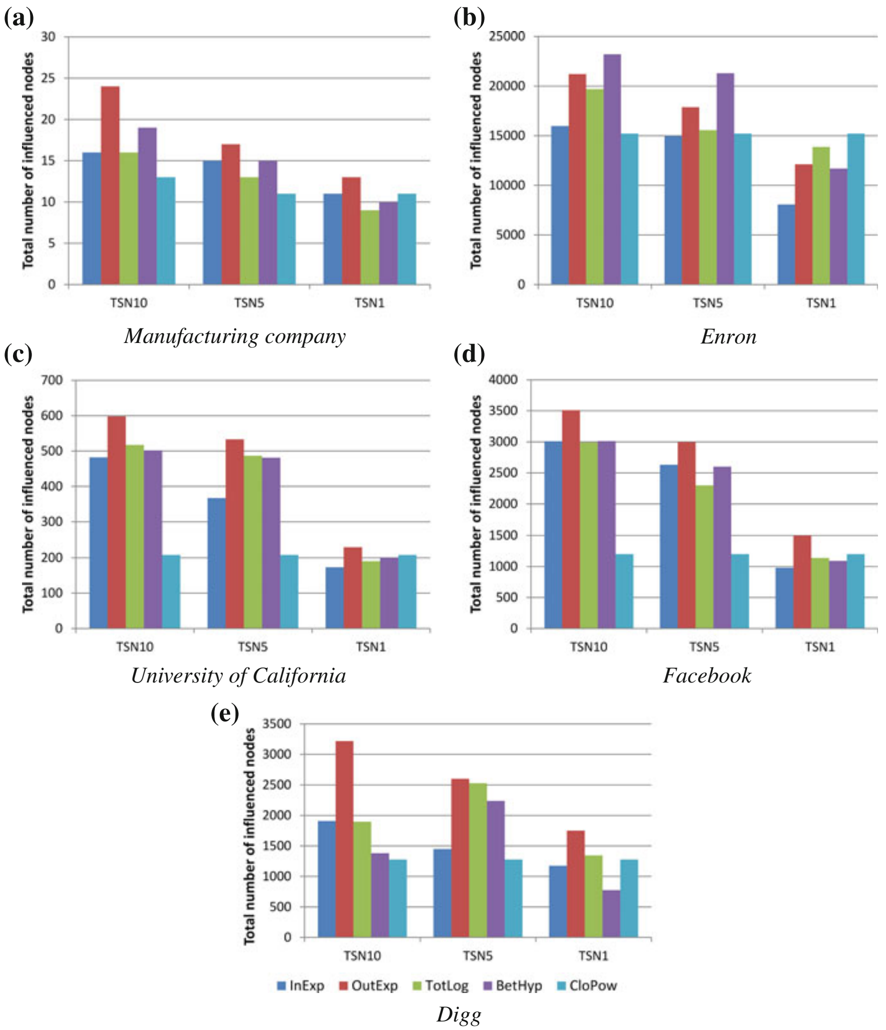


Fig. 14.7 The total number of influenced nodes for all networks and structural measures used for seeding as well as for different datasets, the threshold level $\theta = 0.75$. **a** Manufacturing company. **b** Enron **c** University of California **d** Facebook **e** Digg

14.6.3 Results

Results revealed that indeed for the aggregated (static) network, i.e. TSN^1, the total number of the influenced nodes is the lowest (the right group of bars in Fig. 14.7) and the best performing network type is the one with the biggest number of time windows, i.e. TSN^{10}—the left hand side group of bars, Fig. 14.7. Overall, the final number of influenced nodes for the 10-windows networks (TSN^{10}) was about double as much as for a single network TSN^1, see Fig. 14.7. In this figure the first part of the heuristics name is the measure type, i.e. In—in-degree, Out—out-degree, Tot—total degree, Bet—betweenness, Clo—closeness. The second part of the name represents the type of forgetting—exponential (Exp), logarithmic (Log), hyperbolic (Hyp) or power (Pow), for details see [56].

It confirms the initial hypothesis that using dynamic network it is possible to better utilize the information in original data and finally select better seeds. When about introduced measures, the ones based on forgetting properties outperformed others.

What is more, the greater granularity, the better chance to choose the proper seeds, especially if taking time into consideration by means of time-dependent measures, such as based on linear forgetting. When trying to explain this phenomenon, once again the intuition is suggesting that the increasing granularity is helpful in terms of better representation of the network dynamics, so the sensitivity of the introduced measures increases—they reflect dynamics to a greater extent.

14.6.4 Summary

The above presented experiment shows that indeed the temporal aspects of networks are helpful in building seed strategies and that this direction should be further exploited to verify the initially confirmed assumption that the network dynamics helps in maximizing the spread of influence. In the next section it is shown how researchers try to make use of this direction by presenting recent advancements. As the literature overview shows, the problem here is being solved in different ways, not only as in the presented experiment, but all of these have something in common—they look at the history to maximize the spread.

14.7 Maximizing the Spread of Influence in Dynamic Networks

This section is devoted to presenting recent advancements in maximizing the spread of influence in dynamic networks which are definitely closer to real-world setting. Since most of the research presented here makes the approaches more data-dependent, reader has to have in mind that in contrast to purely structural algorithms described in Sect. 14.4, the application of the approaches introduced below may be limited, since not always the researchers have the full information about a social networks (e.g. communication content). Of course it is not an argument to avoid this direction, but just a loose remark.

14.7.1 Learning Influence Probabilities

Before thinking about maximizing the spread in real-world dynamic networks, there should be at least one limitation overcome, which is how to assign the influence probabilities making them more aligned to the reality. As it was presented in Sect. 14.2, the influence is often drawn from a distribution or hard-wired. When thinking of real-world scenarios, this assumption makes the results of modelling spread of influence questionable. In this area there are just few papers which try to deal with this problem.

The research on this topic started with the work of Tang et al. [70]. Here, the researchers try to avoid learning influence probabilities from the network position of a node, since they assume that different peers of a node may have different influence on it. For instance, our friends may be more influential in the area of private live (trends, friends etc.), while relatives from work may have stronger influence in company-related topics. To take this into account, the authors decided to analyse the content of the communication to build a model of *Topical Affinity Propagation* (TAP). This approach tries to assign influence values over edges between nodes which are topic-specific. So in this case a node may have multiple edges with its neighbour and each of them represents different topic altogether with different influence weights. The authors base on a concept of *factor graphs* [45], in which the observation data are cohesive on both local attributes and relationships. Moreover, to make the approach scalable, they do the following: define a *Topical Factor Graph* (TFG), then they introduce *Topical Affinity Propagation* and finally they try to make the approach scalable for large networks, either by using Map-Reduce approach or a parallel update rule. Their main goal of the proposed idea is expert identification, but this approach suits well for just learning real influence probabilities, as the real-datasets experiments show.

Another approach in this area was proposed by Saito et al. in [65]. Here, the authors focus on the IC model and they base on a likelihood of so-called *episodes*, which are in fact nodes that became influenced in consecutive time-windows. Then they compare neighbouring episodes—the one in time t, which is $D(t)$ in authors' notation, and the next one in $t+1$ (defined as $D(t+1)$) to see whether the neighbouring nodes were in $D(t)$ and $D(t+1)$. If so, it is probable that the newly influenced node v_z from $D(t+1)$ was influenced by v_y from $D(t)$ in time $t+1$. It is possible, because the IC model gives just a single chance to a node to influence its neighbours, so the influence may happen only shortly after the node itself becomes influenced. Then the researchers use the expectation maximization technique to obtain the values of likelihood functions of θ. Experiments conducted on a blogging platform confirm that this approach may be right in terms of obtaining the influence probabilities by learning from past data.

The last work presented here is the work of Goyal et al. [26]. In contrast to the previous work the authors generalize their approach to every influence model following the submodularity property, which makes it more universal (e.g. covering LT and IC). As the reader may remember, the submodularity was the property which allowed to introduce many improvements in the area of maximizing the spread of influence, see Sect. 14.4. In this work the researchers base on two sources: the temporal network and an action log which represents the activity of users. In detail, the action log is defined as a relation containing tuples (u, a, t_u), where u represents a node $\in V$, a—an action from *the universe of actions* and t_u the time when the user u performed the action a. By proposing models for capturing static and dynamic influence the authors are able to compute the probabilities of influence and they reduce the number of scans of typically-huge action log. Moreover, they are able to predict at which time a user will take an action.

14.7.2 Real-World Datasets Evaluation

An important set of hints of how to seek for influential individuals is obtained from the analysis of real-world datasets. In here the case is not to learn influence probabilities to be applied for artificial models, but to get the understanding of what really matters by analysing the influence paths. This information is helpful in building real-world strategies and supplements the mathematical approaches by the knowledge on how influence works in variety of social networks. The only drawback of such analysis is that it is mostly data-dependent, since it emphasizes attributes of users or networks which cannot be generalized easily.

One of the most interesting works in this area is [15]. Authors analyse Twitter social network in order to find the most important factors of individuals that have the greatest influence. Results revealed that the in-degree measure is not a good influence indicator, at least for Twitter. Users with high in-degree not necessarily are the ones that also have the significant influence on others. Moreover an interesting conclusion is drawn that individuals are influential across many topics, i.e. the influence is rarely topic-dependent, but node-dependent. Lastly it was shown that nodes cannot build their influence instantly, it is rather a long-lasting process of becoming important in the neighbourhood. This conclusion suggests that it is barely impossible to insert into the network individuals that become influential fast. Instead the marketing strategies should focus on finding influential nodes and convincing them to opt for a product or service. Some other research on looking for influential nodes on Twitter is presented in [3, 75].

The next study uses Epinions.com portal dataset to find out who is responsible for the most influential reviews of products. By using text-mining techniques the influential power of real online users through their reviews was calculated and combined with the RFM solution which tracks users past behaviour [35]. It was observed that users contributing the most are not necessarily perceived as influential ones, i.e. the experience based on the number of written reviews is not the most important. Secondly, the most important reviewers wrote the reviews in a very emotional manner, i.e. they put a lot of effort to make them sound very subjective instead of objective. It means that neutral reviews are not considered as valuable by others.

It is observed that the approach of finding influential nodes corresponds well to obtaining the influence probabilities, since in the former strategy it is possible to adjust close-to-real influence probabilities, while the latter shows which factors may be additionally used as the ones for choosing the seed set.

14.7.3 Social Influence Maximization

The approach which combines the action log and a social graph presented by Goyal et al. in [26] was later extended to maximize the spread of influence in [27]. Firstly, the authors show that basing on real-world data (action logs or history of past propagations) is crucial, since only by knowing real influence probabilities any algorithm for influence maximization may be accurate. So the initial assumption is that these probabilities have to be computed by real-world data. Then they propose so-called *Credit Distribution model* (CD) which bases on different assumptions than LT and IC models, since it considers actions as a source of influence in network. Authors introduce propagation graphs which include nodes that were neighbours in graph E and performed the same action but in different time. Here, a node performing an action earlier may be considered as a potential influencer of its neighbours taking this action later. So, in fact, the initial graph is static, but the actions introduce the dynamics here. Under the credit model for each action performed by a node, all nodes that took this action earlier and are neighboured to this node receive credits for being potential influencers, and this is a recursive operation. Then the nodes that maximize the influence in the whole network under so-defined model offering $(1 - \frac{1}{e})$—approximation comparing to the optimal solution are being found keeping the scalability as well. The biggest achievement here is the lack of need to perform costly Monte Carlo simulations, but it is because of different model definition. However, the results show that the CD model and the method to choose seeds allow to outperform common approaches for LT and IC influence maximization offering also speed improvement. It is worth to study this paper also because the authors compare the seed sets provided by different models.

In a work of Mathioudakis et al. [54] the researchers use past propagation log and a social graph to find k most influential links, i.e. links that will maximize the propagation. However, what they do is making significant reduction in the search space by benefiting from sparsification. They apply their approach to IC model and propose a *Spine*, dynamic programming algorithm which propose a significant improvement in speed offering accuracy close to optimal. Spine is structured in two phases. During the first phase it selects a set of arcs D_0 that yields a log-likelihood larger than $-\infty$. This is done by means of a greedy approximation algorithm for the Hitting Set NP-hard problem. During the second phase, it greedily seeks a solution of maximum log-likelihood, i.e. at each step the arc that offers the largest increase in log-likelihood is added to the solution set [9].

An interesting approach of considering time-varying influences is proposed in [50] where authors consider the delayed influence process, i.e. the influence of a node to its neighbours may vary in time. It places the problem closer to reality, where people take more care to recent incidents rather than to the older ones. The authors propose *Influence Spreading Paths* as a method of measuring the influence of a node, i.e. $ISP(u, S)$ represents all spreading paths that end with user u. By using them the authors compute the activation probability of a user u and thanks to that they are able to find seeds faster than by using greedy algorithm. So the time factor incorporated here is not the time reflecting the dynamics of the network but the changes in influence probabilities. However, the researchers of this chapter indicate that this is another way of representing the network dynamics and as such it can be used for solving the problem in a dynamic environment.

The problem of influence maximization in dynamic social networks was just recently explicitly stated in [1]. As the authors claim, to their knowledge they propose the first set for time-sensitive methods for influence maximization. In this work researchers use the transmission matrix which contains the time-dependent functions for influence spread to find solution for two separate problems. Firstly, they would like to pick k nodes at time t_1 to maximize the influence at time t_2—this problem lies closer to classic influence maximization problem, but it incorporates the time factor. Secondly, when observing the influence spread at time t_2 they would like to know which nodes most probably were responsible for the influence spread at time t_1. To deal with these problems, they introduce *Backward and Forward Influence Algorithms*. When looking at influence maximization problem, authors try to solve it similarly to the greedy algorithm, but now each iteration means another time-step. In conducted experiments it is shown that the time-dependent solutions outperformed the static ones showing that this direction should be further exploited.

To two more works in this area which tackle the problem differently are worth referring. In [47] researchers try to find successor nodes for removed seeds. It is a relatively different research question than in a typical influence maximization problem, since now the budget k will increase, but this approach incorporates the dynamics of networks showing that considering it is crucial. In [36] authors try to take into account the availability factor of nodes, which indeed is the embedded dynamics of the network, trying to improve the overall influence spread in networks. Again, experiments' results confirm that this direction helps the seeding strategies in obtaining better results.

The idea of maximizing spread of influence in dynamic networks is a relatively new one, but as the above literature review shows, it seems that considering the dynamics of networks in the influence process is important and already some solutions are being proposed. However, the work in this area has just begun and we should expect some improvements shortly. One of the reasons is that the dynamics in social networks is something natural rather than unusual and it is already agreed that the influence maximization problem should be considered in this real-world setup.

14.8 Summary

In this work the problem of influence maximization in social networks was presented. By starting with the most typical models of social influence, namely the Linear Threshold, Independent Cascade, Voter Model and the Naming Game the influence maximization for social networks was defined. After that it was shown how the solutions for it developed for the static case. However, as the static case barely fits the complex reality of dynamic social networks which are definitely more often found in real-world, a short introduction to temporal networks was presented just after. To confirm that the temporal solutions may be helpful in seeding strategies, a short experimental study was quoted and discussed. In the last part of this chapter the recent developments in the area of maximizing the spread of influence in dynamic networks were shown.

However, the problem lies in how to put all the information provided here into a synthetic form to conduct a successful marketing campaign. It seems that apart from theoretical approaches a great deal of real-world dataset evaluation suggestions has to be included in order to find out who is considered as an influential person in particular social network. If there is only a pure network structure available, designers of campaign have no other option then use purely structural approaches. But when attributes of individuals or time-sensitive data are obtained, the seed selection process should use dataset-dependent information. An interesting case would be a comparison of how data-dependent seeding strategies perform against structural ones. Yet it seems that ignoring the temporal information may be one of the worst strategies.

The idea of this work was to show how the problem developed and was solved in a chronological order and what are the current challenges in this area. Since the apparatus for temporal networks is already established [34] and the problem for the dynamic case is clearly stated [1], nothing should stop other researchers in this area from developing new algorithms for the real-world scenarios.

Acknowledgments This work was partially supported by the fellowship co-financed by the European Union within the European Social Fund, by the European Commission under the 7th Framework Programme, Coordination and Support Action, Grant Agreement Number 316097, ENGINE—European research centre of Network intelliGence for INnovation Enhancement (http://engine.pwr.wroc.pl) and by The National Science Centre, the decision no. DEC-2013/09/B/ST6/02317. The calculations were carried out in Wroclaw Centre for Networking and Supercomputing (http://www.wcss.wroc.pl), grant No 177.

References

1. Aggarwal, C.C., Lin, S., Philip, S.Y.: On influential node discovery in dynamic social networks. In: SDM, pp. 636–647. SIAM, Anaheim (2012)
2. Aral, S., Walker, D.: Identifying influential and susceptible members of social networks. Science **337**(6092), 337–341 (2012)

3. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone's an influencer: quantifying influence on twitter. In: Proceedings of the fourth ACM international conference on Web search and data mining, pp. 65–74. ACM (2011)
4. Barabási, A.L.: Bursts: The Hidden Patterns Behind Everything we do, From Your E-mail to Bloody Crusades. Penguin (2010)
5. Barbieri, N., Bonchi, F., Manco, G.: Topic-aware social influence propagation models. Knowl. Inf. Syst. **37**(3), 555–584 (2013)
6. Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A.: The architecture of complex weighted networks. Proc. Natl. Acad. Sci. U.S.A. **101**(11), 3747–3752 (2004)
7. Berger, E.: Dynamic monopolies of constant size. J. Comb. Theor. Ser. B **83**(2), 191–200 (2001)
8. Bharathi, S., Kempe, D., Salek, M.: Competitive influence maximization in social networks. In: Internet and Network Economics, pp. 306–311. Springer, Heidelberg (2007)
9. Bonchi, F.: Influence propagation in social networks: a data mining perspective. IEEE Intell. Inf. Bull. **12**(1), 8–16 (2011)
10. Borodin, A., Filmus, Y., Oren, J.: Threshold models for competitive influence in social networks. In: Internet and Network Economics, pp. 539–550. Springer, Heidelberg (2010)
11. Bródka, P., Saganowski, S., Kazienko, P.: Ged: the method for group evolution discovery in social networks. Soc. Netw. Anal. Min. **3**(1), 1–14 (2013)
12. Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y., Shir, E.: A model of internet topology using k-shell decomposition. Proc. Natl. Acad. Sci. **104**(27), 11150–11154 (2007)
13. Carrington, P.J., Scott, J., Wasserman, S.: Models and Methods in Social Network Analysis. Cambridge University Press, Cambridge (2005)
14. Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J.F., Vespignani, A.: Dynamics of person-to-person interactions from distributed RFID sensor networks. PloS ONE **5**(7), e11596 (2010)
15. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring user influence in twitter: the million follower fallacy. ICWSM **10**, 10–17 (2010)
16. Chen, W., Lu, W., Zhang, N.: Time-critical influence maximization in social networks with time-delayed diffusion process. arXiv:1204.3074 (2012)
17. Chen, W., Wang, C., Wang, Y.: Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1029–1038. ACM (2010)
18. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 199–208. ACM (2009)
19. Clifford, P., Sudbury, A.: A model for spatial conflict. Biometrika **60**(3), 581–588 (1973)
20. DallAsta, L., Baronchelli, A., Barrat, A., Loreto, V.: Nonequilibrium dynamics of language games on complex networks. Phys. Rev. E **74**(3), 036105 (2006)
21. De Choudhury, M., Sundaram, H., John, A., Seligmann, D.D.: Social synchrony: predicting mimicry of user actions in online social media. In: Computational Science and Engineering, 2009. CSE'09. International Conference on, vol. 4, pp. 151–158. IEEE (2009)
22. Domingos, P., Richardson, M.: Mining the network value of customers. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 57–66. ACM (2001)
23. Durrett, R., Durrett, R., Durrett, R., Durrett, R.: Lecture notes on particle systems and percolation. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA (1988)
24. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. Pers. Ubiquitous Comput. **10**(4), 255–268 (2006)
25. Goldenberg, J., Libai, B., Muller, E.: Talk of the network: a complex systems look at the underlying process of word-of-mouth. Mark. Lett. **12**(3), 211–223 (2001)
26. Goyal, A., Bonchi, F., Lakshmanan, L.V.: Learning influence probabilities in social networks. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 241–250. ACM (2010)

27. Goyal, A., Bonchi, F., Lakshmanan, L.V.: A data-based approach to social influence maximization. Proc. VLDB Endowment **5**(1), 73–84 (2011)
28. Goyal, A., Bonchi, F., Lakshmanan, L.V., Venkatasubramanian, S.: On minimizing budget and time in influence propagation over social networks. Soc. Netw. Anal. Min. **3**(2), 179–192 (2013)
29. Goyal, A., Lu, W., Lakshmanan, L.V.: Celf++: optimizing the greedy algorithm for influence maximization in social networks. In: Proceedings of the 20th International Conference Companion on World wide web, pp. 47–48. ACM (2011)
30. Goyal, A., Lu, W., Lakshmanan, L.V.: Simpath: an efficient algorithm for influence maximization under the linear threshold model. In: Data Mining (ICDM), 2011 IEEE 11th International Conference on, pp. 211–220. IEEE (2011)
31. Granovetter, M.: Threshold models of collective behavior. Am. J. Sociol. **83**(6), 1420 (1978)
32. Hedström, P., Bearman, P.: The Oxford Handbook of Analytical Sociology. Oxford University Press, Oxford (2009)
33. Holley, R.A., Liggett, T.M.: Ergodic theorems for weakly interacting infinite systems and the voter model. The Annals of Probability, pp. 643–663 (1975)
34. Holme, P., Saramäki, J.: Temporal networks. Phys. Rep. **519**(3), 97–125 (2012)
35. Hughes, A.M.: Strategic Database Marketing. McGraw-Hill, New York (2006)
36. Jankowski, J., Michalski, R., Kazienko, P.: Compensatory seeding in networks with varying avaliability of nodes. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 1242–1249. ACM (2013)
37. Javarone, M.A.: Social influences in the voter model: the role of conformity. arXiv preprint arXiv:1401.0839 (2014)
38. Jiang, Q., Song, G., Cong, G., Wang, Y., Si, W., Xie, K.: Simulated annealing based influence maximization in social networks. In: AAAI (2011)
39. 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, pp. 137–146. ACM (2003)
40. Kempe, D., Kleinberg, J., Tardos, É.: Influential nodes in a diffusion model for social networks. In: Automata, Languages and Programming, pp. 1127–1138. Springer, Heidelberg (2005)
41. Kindermann, R., Snell, J.L., et al.: Markov Random Fields and their Applications, vol. 1. American Mathematical Society Providence, R.I. (1980)
42. Klimt, B., Yang, Y.: The enron corpus: a new dataset for email classification research. In: Machine learning: ECML 2004, pp. 217–226. Springer, Heidelberg (2004)
43. Król, D.: On modelling social propagation phenomenon. In: N. Nguyen, B. Attachoo, B. Trawiński, K. Somboonviwat (eds.) Intelligent Information and Database Systems, Lecture Notes in Computer Science, vol. 8398, pp. 227–236. Springer, Heidelberg (2014)
44. Król, D.: Propagation phenomenon in complex networks: theory and practice. New Gener. Comput. **32**(3–4), 187–192 (2014)
45. Kschischang, F.R., Frey, B.J., Loeliger, H.A.: Factor graphs and the sum-product algorithm. Inf. Theor. IEEE Trans. **47**(2), 498–519 (2001)
46. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 420–429. ACM (2007)
47. Li, C.T., Hsieh, H.P., Lin, S.D., Shan, M.K.: Finding influential seed successors in social networks. In: Proceedings of the 21st International Conference Companion on World Wide Web, pp. 557–558. ACM (2012)
48. Li, Y., Chen, W., Wang, Y., Zhang, Z.L.: Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, pp. 657–666. ACM (2013)
49. Liggett, T.M.: Interacting Particle Systems. Springer, Berlin (1985)
50. Liu, B., Cong, G., Xu, D., Zeng, Y.: Time constrained influence maximization in social networks. In: ICDM, pp. 439–448 (2012)

51. Lu, Q., Korniss, G., Szymanski, B.K.: The naming game in social networks: community formation and consensus engineering. J. Econ. Interac. Coord. **4**(2), 221–235 (2009)
52. Maity, S.K., Mukherjee, A., Tria, F., Loreto, V.: Emergence of fast agreement in an overhearing population: the case of the naming game. EPL (Europhysics Letters) **101**(6), 68,004 (2013)
53. Masuda, N., Holme, P.: Predicting and controlling infectious disease epidemics using temporal networks. F1000Prime Reports 5, 6 (2013)
54. Mathioudakis, M., Bonchi, F., Castillo, C., Gionis, A., Ukkonen, A.: Sparsification of influence networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 529–537. ACM (2011)
55. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1953)
56. Michalski, R., Kajdanowicz, T., Bródka, P., Kazienko, P.: Seed selection for spread of influence in social networks: Temporal versus static approach. New Gener. Comput. (2014) (in press)
57. Michalski, R., Palus, S., Kazienko, P.: Matching organizational structure and social network extracted from email communication. In: Business Information Systems, pp. 197–206. Springer, Berlin (2011)
58. Mobilia, M.: Commitment versus persuasion in the three-party constrained voter model. J. Stat. Phys. **151**(1–2), 69–91 (2013)
59. Opsahl, T., Panzarasa, P.: Clustering in weighted networks. Soc. Netw. **31**(2), 155–163 (2009)
60. Pathak, N., Banerjee, A., Srivastava, J.: A generalized linear threshold model for multiple cascades. In: 2010 IEEE 10th International Conference on, Data Mining (ICDM), pp. 965–970. IEEE (2010)
61. Peleg, D.: Local majority voting, small coalitions and controlling monopolies in graphs: a review. In: Proceedings of 3rd Colloquium on Structural Information and Communication Complexity, pp. 152–169 (1997)
62. Pfitzner, R., Scholtes, I., Garas, A., Tessone, C.J., Schweitzer, F.: Betweenness preference: quantifying correlations in the topological dynamics of temporal networks. Phys. Rev. Lett. **110**(19), 198,701 (2013)
63. 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, pp. 61–70. ACM (2002)
64. Rogers, T., Gross, T.: Consensus time and conformity in the adaptive voter model. Phys. Rev. E **88**(3), 030,102 (2013)
65. Saito, K., Nakano, R., Kimura, M.: Prediction of information diffusion probabilities for independent cascade model. In: Knowledge-Based Intelligent Information and Engineering Systems, pp. 67–75. Springer, Berlin (2008)
66. Schelling, T.: Micromotives and Macrobehavior. WW Norton and Company, New York (1978)
67. Schrijver, A.: Combinatorial Optimization: Polyhedra and Efficiency, vol. 24. Springer, Berlin (2003)
68. Shakarian, P., Paulo, D.: Large social networks can be targeted for viral marketing with small seed sets. In: Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pp. 1–8. IEEE Computer Society, Canada (2012)
69. Sun, J., Tang, J.: A survey of models and algorithms for social influence analysis. In: Social Network Data Analytics, pp. 177–214. Springer, New York (2011)
70. 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, pp. 807–816. ACM (2009)
71. Vanhems, P., Barrat, A., Cattuto, C., Pinton, J.F., Khanafer, N., Régis, C., Kim, B.a., Comte, B., Voirin, N.: Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PloS ONE **8**(9), e73970 (2013)
72. Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P.: On the evolution of user interaction in facebook. In: Proceedings of the 2nd ACM Workshop on Online Social Networks, pp. 37–42. ACM (2009)

73. Xie, J., Emenheiser, J., Kirby, M., Sreenivasan, S., Szymanski, B.K., Korniss, G.: Evolution of opinions on social networks in the presence of competing committed groups. PloS ONE **7**(3), e33215 (2012)
74. Xie, J., Sreenivasan, S., Korniss, G., Zhang, W., Lim, C., Szymanski, B.K.: Social consensus through the influence of committed minorities. Phys. Rev. E **84**(1), 011,130 (2011)
75. Ye, S., Wu, S.F.: Measuring Message Propagation and Social Influence on Twitter.com. Springer, Heidelberg (2010)
76. Zafarani, R., Abbasi, M.A., Liu, H.: Social Media Mining: An Introduction. Cambridge University Press, Cambridge (2014)
77. Zhang, W., Lim, C., Korniss, G., Szymanski, B.: Spatial Propagation of Opinion Dynamics: Naming Game on Random Geographic Graph. arXiv:1401.0115 (2013)