International Journal of Scientific Research in Computer Science, Engineering and Information Technology
 | Volume 3 | Issue 1 | ISSN : 2456-3307

A Novel Approach to Limit the Spread of Wrong Information in Social Networks

K. Ravikumar[1], N. Sindhuja[2]

[1]Assistant. professor Department of Computer Science, Tamil University, Thanjavur, Tamil Nadu, India

[2]Research Scholar, Department of Computer Science, Tamil University, Thanjavur, Tamil Nadu, India

ABSTRACT

In this effort, we study the idea of competing campaigns in a social network. By demonstrating the spread of effect in the presence of competing campaigns, we provide necessary tools for applications such as emergency response where the goal is to limit the spread of misinformation. We revisited the problematic of effect limitation where a bad campaign twitches spreading from a certain node in the grid and use the notion of limiting campaigns to counteract the misinformation. The problem can be summarized as identifying a subset of folks that need to be convinced to adopt the competing (or \decent") movement so as to minimalize the number of people that adopt the \bad" campaign at the end of both propagation processes. We demonstration that this optimization problematic is NP-hard and deliver estimate assurances for an avaricious answer for various meanings of this problem by proving that they are submodular. Though the greedy algorithm is a polynomial time algorithm, for today's big scale social networks even this answer is computation-ally actual expensive. Consequently, we study the presentation of the degree importance experiential as well as other heuristics that have insinuations on our exact problem. The experiments on a number of close-knit regional networks obtained from the Facebook social network show that in most belongings inexpensive heuristics do in fact compare well with the greedy approach.

Keywords : Social Systems, Evidence Cascades, Misrepresentation Proliferation, Competing Movements, Submodular Occupations

I. INTRODUCTION

Until very lately, information about the vast mainstream of public events has been provided by, or littered through, the mass media which had almost complete independence over the decisions as to which piece of data is newsworthy. This few-to-many information model has been shat-treed by advances in technology during the last decade, especially with the adoption of the online social networks [11]. Social networks have remained shown to have benefits as a medium for fast, extensive information dissemination. They provide fast contact to large scale news data, occasionally even before the mass media as in the case of statement of death of Michael Jackson. They also serve as an intermediate to together achieve a social goal. For instance, with the use of collection and event sheets in Facebook, events such as Day of Action disputes reached thousands of protestors [12].

Though the comfort of information propagation in social net-works can be very helpful, it can also have disruptive effects. One such example was observed during the recent shootings at Fort Hood, Texas, when a soldier inside the base sent out messages via

Twitter as the event unfolded. This woman improper reports of multiple shooters and shelling sites quickly spread over the social network and even to the mass media where it was stated on television programs [16]. Additional sample is the binge of misrepresentation on swine u in Twitter. The blowout of misinformation in this case stretched a very large scale causing panic in the people. By way of folks were being misinformed on the issue, they also contributed in this misinformation tendency by repeating it and therefore distributing it even further. Though social networks like Twitter are the main basis of news for many persons today, they are still not careful reliable due to such glitches.

Obviously, instruction for social networks to serve as a dependable platform for distributing critical information, it is essential to have tools to bind the effect of distortion. In this revision one of our foremost objectives is to address this specific problem. In the presence of a distortion flow, we aim to end the most optimal way of disseminating good information that will minimalize the overwhelming belongings of the distortion campaign. On behalf of example in the case of [17, 16], we pursue ways of making sure that most of the users of the social network hear about the correct information before the bad one. In this method, we can make communal networks a more trustworthy or dependable source of information. In addition to the implication our work has in limiting the effect of distortion, the methods we introduce can also be practical to any two competing campaigns that are concurrently spreading through the network. Since in a real social network, here are usually binary or more correlated information forces happening instantaneously, we believe capturing this typical is crucial to getting a more truthful classical of real social networks.

In this exertion, we study the problematic of diminishing number of people that accept the misrepresentation and confirm that level however the general problem does not exhibit the submodular property, certain limited versions of it are in fact submodular. It exploits this property to deliver well-organized solutions with approximation bounds. We also assess the performance of our algorithm on a number of close area networks found from the Facebook social network comparing its performance with some well-known heuristics such as grade consequence as well other heuristics and display that in many cases heuristics achieve similar to the computationally additional penetrating greedy method.

II. RELATED WORK

The documentation of operators or view leaders in a social network is a problematic that has conventional an important quantity of care in recent research. In the effect maximization problem, given a probabilistic model of data dispersal such as Autonomous Cascade Model, a network graph, and a economical k, the objective is to select a set A of size k for initial activation so that the expected value of f(A) (size of cascade created by choosing set A) is maximized [8]. By a well-organized, healthy solution to this problem, it remains possible to extensively distribute significant data in a social network. Early mechanisms trusted on heuristics such as bulge degree formerly coldness importance to select the set A. Though the problematic of finish an best solution in this perfect is NP-hard, it has been exposed that there is a greedy algorithm that yields a spread that is within 1 1=e of best [17]. This solution depends on Monte Carlo simulations which are computationally intense.

Work has been done on improving the presentation of greedy algorithms for Influence maximization [5, 19], but scalability remains a significant challenge. In adding to the measure issues that remain inherently there, these definitions of influential users disregard certain features of the real communal networks such by way of the existence of competing movements. In this work we reflect different models of communication that incorporate dissimilar aspects of real social networks. The works that are closest to the one introduced in this paper are [17, 22]. Similar to those works, we categorize a problem in a social network that involves classifying influential nodes and education the possibility of a solution to this problem. However, our problematic preparation is more general in that, we perfect the existence of rival cascades dispelling in a system.

The being of rival campaigns has been seized by a number of studies recently. Dubey et al. [9] study the badly-behaved as a web game focusing on quasi-linear model and reflect various cost, bene t and externalities functions for competing rms. They study the existence of Nash Equilibrium (NE) and demonstration that NE is unique if there is enough competition between rams or if their valuations of clients are anonymous. Bharathi et al. [3] augment the Offensive Cascade Model to imprisonment the existence of competing campaigns in a network. Their diffusion model is parallel to the unique studied in our effort and imprisonments the judgement subjects that are crucial toward competing movement's optimization problems. The algorithmic problematic they method is: Given that there is more than one movement dissipating in a net-work and each campaign can select a set of initial adopters so as toward maximize their bene t, i.e. number of people accepting their product, what is the best strategy for the players? This work educations the problem from both the rest and last player's viewpoints and shows that the problem of choosing the early adopters for the last actor is submodular.

They also introduce a FPTAS for the rest player when the network structure is a tree. Carnes et al. [4] contemplate the same delinquent from the last player's perspective and use one diffusion model where nodes of the network choose the campaign to adopt w.r.t. their distance to the early adopters of the campaigns and another model where the nodes make a uniform random choice among its active neighbors. They present experimental results that show that the greedy approach with the approximation bounds performs better than the heuristics but the difference is not significant which agrees with the results presented in this work.

They also experimentally show that the best strategy for the rest player is to choose high degree nodes. Kostka et al. study opposing movements also as a game hypothetical problem and show that being the rest player, i.e. the rest to decide, is not always advantageous. Both [4, 3] use diffusion models where the competing campaigns propagate exactly the same way, i.e. the probability of diffusion on a certain edge is the same for all campaigns and all campaigns start at the same time. In our effort, we revision the circumstance where the rival campaigns have dissimilar acceptance rates and one remains in response to the other, and so campaign of the last player remains started with a delay.

III. DIFFUSION OF MISINFORMATION

A social network can be demonstrated as a directed graph G = (N; V) containing of nodes N and edges V. A node v is a neighbor of w if and only if there is an edge from w to v in G. In the setting of influence spread, the set of nodes, N can be viewed as the operators of the social network. If a user w is a \friend" of additional user v, there is a direct announcement link, an edge $e_{v;w}$ in G. In addition to this, allocate a weight $p_{v;w}$ to each edge $e_{v;w}$ which is used to perfect the direct influence v has on w or conversely the likelihood that v will forward certain information it obtains to its neighbor w. Note that in this setting, \friendship" is an asymmetric association which enables us to model the case where the inspiration one user has on a friend is different than the effect this friend has on that user.

3.1 Diffusion Models

Self-governing cascade is one of the most basic and well-studied diffusion models that has been used in different con-texts [10, 23, 13, 14]. In Self-governing

Cascade Model, the technique starts with an initial set of active nodes A_0, and the process unfolds in discrete steps according to the following randomized rule. When node v first developments energetic in step t, it is given a single chance to stimulate each currently sedentary neighbor w; it succeeds with a probability $p_{v;w}$ in-dependent of the history thus far. If v prospers, then w will become active in step t + 1; but whether or not v succeeds, it cannot make any further attempts to stimulate w in sub-sequent rounds. The process runs until no additional activations are conceivable. If w has manifold newly activated neighbors, their efforts are sequenced in an uninformed order.

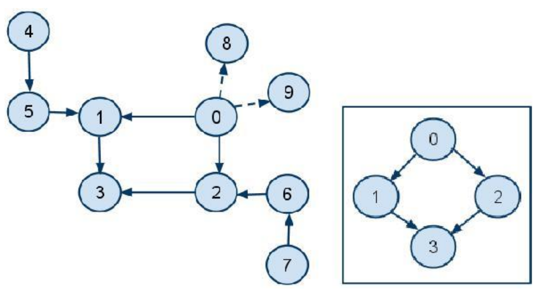
The Multi-Campaign Independent Cascade (MCICM) model. Here models the diffusion of two cascades evolving concurrently in a system. Let C (stands for \campaign") and L (stands for \limiting campaign") denote the two cascades and the initial set of active swellings for cascade L is denoted by A_L. Similarly, A_C signifies the initial set of vigorous nodes in C. The process unfolds again in separate time steps. When a node v first produces active in campaign L (or C) in step t, it is given a single chance to stimulate each currently sedentary neighbor w in campaign L (or C) and it succeeds with likelihood $p_{L;v;w}$ (or $p_{C;v;w}$) given that no national of w tries activating w in the competing campaign at the same step. It also refer to $p_{L;v;w}$ (or $p_{C;v;w}$) as the probability of the edge $e_{v;w}$ being live. If nearby are two or more nodes trying to stimulate w at a given time step, at most one of them can succeed. In self-determining cascade, when w has recurrent newly started nationals, their attempts are sequenced in arbitrary order. However, in our studies, will shoulder that there is a natural order to the two campaigns, more specifically one is \good" while the other is the \bad" campaign and if the \bad information" and the \good information" reach a node w at the same step, \good information" takes effect. Once a node grows active in one campaign, it never becomes inactive or fluctuations movements and the process endures until there is no newly activated node in either campaign.

It also considers another model of dissemination in which the probabilities of each edge being live is self-governing of the campaign. In this background it only associate one likelihood $p_{v;w}$ with each edge $e_{v;w}$ and hence the model becomes al-most identical to the Independent Cascade Model. No matter which evidence reaches a node v, v forwards this information to its neighbor w with likelihood $p_{v;w}$. Although this model is not a perfect t for immunization of misrepresentation (since nodes of the network would be more willing to share \good" information), it is a good t for modeling challenging campaigns where the two information waterfalls are more likely to be of similar \quality" and the nodes would agree to the campaign that reaches out to them first. Consider for example two articles L and C about the same event dispersal through a social network. The probability of a user advancing article L and C is more dependent on the news itself rather than which activity the news is from. Comparable to the Multi-Campaign Independent Cascade model, there are three conditions a node can be in; inactive, in campaign L, in campaign C and once a node becomes active in either L or C, it cannot modification its state again. As beforehand, it assumes that in the case of immediate trials of activation at a node, movement L is ordered before C. This model Campaign-Oblivious Independent Cascade (COICM). COICM is comparable to the diffusion model used in [3]. How-ever here we assume that one of the movements is arranged over the other one in the case of simultaneous activation trials whereas independent and exponentially distributed continuous random variables are allocated to each edge as delay in [3] to make sure there will be no simultaneous commencement trials. Note that, the algorithms available in Section 4 would also work the dispersal model presented in [3].

3.2 Problem Definition

While a considerable amount of research has been done in the background of influence expansion, a problem that has not received much attention is that of limiting the stimulus of a malicious or incorrect evidence campaign. One strategy to deal with a misrepresentation campaign is to limit the number of users who are willing to accept and spread this misinformation. It assumes that the Multi-Campaign Independent Cascade Model described in Section 3.1 as the prototypical of announcement. Without loss of generalization it will assume that the spread of influence for movement C starts from one node n_a and its existence is illustrious at time step r and at that point the undertaking L is started. It refer to r also as the delay of campaign L. Our aim is to either limit the effect of campaign C or to maximize the effect of L contingent on the specific objective function.

Depending on the background that the influence limitation problem is presented in, we need to consider different objective functions. The objective can be to try and \save" as many persons as possible, to limit the lifetime of the \bad" in-formation movement or to make the most of the effect \good" campaign in the presence of the \bad" campaign. In the next section, we will focus on minimizing the number of people that end up accepting campaign C when information waterfalls from both movements are over. It refer to this problem as the eventual inspiration constraint problem.



(a) A graph representing (b) the shortest spread of campaign C.

Solid path construction for lines characterize the live edges spread and dotted lines presentence for the bad dead edges for the spread movement of information campaign C. Assume that the opponent is node 0. In this case, if there was no contrasting campaign, C would reach A_C = f0, 1, 2,3g

f Given (n_a,r,k) where n_a denotes the adversary and r denotes the time step campaign C is detected and k denotes the number of nodes to initially activate in Lg

Initialize A_L to ; R to 10000

for i = 1 to k do

for each vertex v 2 V A_L do

s_v = 0

for j = 1 to R do

s_v+ = InfLimit(n_a; r; A_L; v)

s_v = s_v=R

//Choose node i that maximizes (A_L [fig) (A_L)

//And set A_L A_L [fig

A_L = A_L [fargmax$_{v2V\ AL}$ fs$_v$gg

Output A_L

The second heuristic consider is early infected. This concept refers to choosing seeds that are predictable to be infected at time step r which well-defined as the delay of movement L. This is corresponding to reaching out to nodes that would be infected early on but after L is started, since those nodes are likely to create a large waterfall for campaign C.

The third heuristic is largest infected. This experiential is very similar to the early infected but rather than simply choosing the nodes that are expected to be infected early on, it aims to choose seeds that are predictable to infect the highest number of people if they were to be infected themselves. In this case restrict ourselves to such nodes that would be infected after time step r. Note that both early diseased and largest infected are more computationally intensive to compute than degree centrality. However, they are still far less intense than the avaricious method that involves shortest path computations. Computation of these heuristics is very comparable to the problem of infection discovery studied in [22] and has been shown to scale to very large networks.

Here appraise how well the avaricious algorithm performs w.r.t. the three heuristics discussed. Note that since inspiration propagation is a stochastic process, we need to complete Monte Carlo Simulations in the order of thousands as part of our algorithm. This is one of the major scalability issues inherent in this type of problematic. However, in our specific problem each simulation involves the expensive computation of straight paths which is critical to eventual influence limitation and this makes ultimate influence constraint even more computationally intense than those of [17, 22]. As part of our experiments, we also evaluated how factors like the degree importance of the adversary, delay of campaign L, and the weight distribution for $p_{C;v;w}$ and $p_{L;v;w}$ influence our excellent of best t algorithm. This requires running thousands of experiments on the same network data. Taking these influences into consideration achieved experiments on 4 provincial network graphs obtained from Facebook.

IV. CONCLUSION

In this effort have thoughtful the algorithmic problem of warning the effects of misrepresentation in a social network. This eventual influence restriction problem. In order to study this problem, rest introduced an announcement model of social networks that joins the notion of correlated campaigns that are distributing simultaneously in a network. We proved that eventual influence restriction problem is NP-hard and therefore an exact solution is in-feasible. We also showed that two variations of this problem on two different announcement models are submodular and therefore a greedy method is definite toward provide a 1= (1 e) approximation. Nonetheless the greedy algorithm is polynomial period, it is still too expensive for today's large scale social networks. Consequently, in addition to the estimate bounds, we also experimentally studied the presentation of the avaricious algorithm, comparing it with 3 different heuristics one of which is degree centrality. We showed that, in many cases, heuristics do comparable to the greedy algorithm, even the simple degree importance heuristic. This may seem counterintuitive at rest glance since it does not adhere too many of the studies that claim poor performance for heuristics such as degree centrality.

Note however that those performance results have been demonstrated on models of diffusion that do not capture the entire reality of social networks, i.e. the fact that there are multiple campaigns spreading concurrently in a network. Notwithstanding the examination claims of poor performance for such heuristics, marketers have been using those heuristics for a very long time with the claim that \it works for them" [12]. This revision provides insights as to why it works in reality. We also recognized the cases where degree position is not a good experimental and presented that in those cases, the largest degree experiential still achieves comparable to the avaricious technique while being computationally less intense. It deliberates unrelated aspects of the problem such as the consequence of initial the limiting campaign early/late, the effect of the properties of the adversary and how prone the populace in general is to longsuffering either one of the arrangements.

V. REFERENCES

[1]. R. M. Anderson and R. M. May. Infectious Diseases of Humans: Dynamics and Control (Oxford Science Publications). Oxford University Press, USA, new ed edition, September 1992.

[2]. N. T. J. Bailey. The mathematical theory of infectious diseases and its applications / [byNorman T. J. Bailey. Gri n, London :, 2nd ed. edition, 1975.

[3]. S. Bharathi, D. Kempe, and M. Salek. Competitive in uence maximization in social networks. In In WINE, pages 306{311, 2007.

[4]. T. Carnes, C. Nagarajan, S. M. Wild, and A. van Zuylen. Maximizing in uence in a competitive social network: a follower's perspective. In ICEC '07: Proceedings of the ninth international conference on Electronic commerce, pages 351{360, New York, NY, USA, 2007. ACM.

[5]. W. Chen, Y.Wang, and S. Yang. E cient in uence maximization in social networks. In Proceedings of the 15th ACM International Conference on Knowledge Discovery and Data Mining, pages 199{208, 2009.

[6]. G. Cornuejols, M. L. Fisher, and G. L. Nemhauser. Location of bank accounts to optimize oat: An analytic study of exact and approximate algorithms. Management Science, 23(8):789{810, 1977.

[7]. O. Diekmann and J. A. P. Heesterbeek. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation (Wiley Series in Mathematical & Computational Biology). Wiley, 1 edition, May 2000.

[8]. P. Domingos and M. Richardson. Mining the network value of customers. In Proceedings of the 7th ACM International Conference on Knowledge Discovery and Data Mining, pages 57{66, 2001.

[9]. P. Dubey, R. Garg, and B. D. Meyer. Competing for customers in a social network. Cowles Foundation Discussion Papers 1591, Cowles Foundation, Yale University, Nov. 2006.

[10]. R. Durrett. Lecture notes on particle systems and percolation. Wadsworth Publishing, 1988.

[11]. Facebook. http://www.facebook.com.

[12]. J. Garrison and C. Knoll. Prop. 8 opponents rally across california to protest gay-marriage ban. Los Angeles Times, November 2008.

[13]. J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters, 2001.

[14]. M. Granovetter. Threshold models of collective behavior. American Journal of Sociology, 1978.

[15]. H. W. Hethcote. The mathematics of infectious diseases. SIAM Rev., 42(4):599{653, 2000.

[16]. K. M. Heussner. Ft. hood soldier causes stir on twitter. ABS News (online), November 2009.

[17]. D. Kempe, J. M. Kleinberg, and E. Tardos. Maximizing the spread of in uence through a social network. In Proceedings of the Ninth ACM International Conference on Knowledge Discovery and Data Mining, pages 137{146, 2003.

[18]. J. O. Kephart and S. R. White. Measuring and modeling computer virus prevalence. In SP '93: Proceedings of the 1993 IEEE Symposium on Security and Privacy, page 2, Washington, DC, USA, 1993. IEEE Computer Society.

[19]. M. Kimura, K. Saito, R. Nakano, and H. Motoda. Social Computing and Behavioral Modeling, chapter Finding In uential Nodes in a Social Network from Information Di usion Data. Springer US, 2009.

[20]. M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identifying in uential spreaders in complex networks. Jan 2010.