Abstract Implementation of Graph Mining Technique using Structural Datum in Viral Marketing

M. Antony Sundarsingh, S.P.Victor

Abstract- Graph mining and marketing has become an important topic of research recently because of numerous applications to a wide variety of business problems in computational biology, chemical data analysis, drug discovery and communication networking. Nowadays Graphs play a vital role everywhere, occupying the social networks and mobile networks to biological net-works and the World Wide Web. Mining big graphs leads too many interesting applications including marketing, news groups, community mining, and many more. In this paper we describe a technique for the implementation of real-time marketing to a Graph Mining pattern. Our findings include designs to survey different aspects of graph mining and management, and provide a compendium for other researchers in the field. The results are revealed for selecting the optimized maximum priority based network selection to implement the marketing action. In the future we will extend our research to propose a Graph-Analysis Implementer for any real-time complex entities.

Keywords- Graph mining, Graph pattern, Graph template, Graph network.

I. INTRODUCTION

A graph is set of nodes, pairs of which might be connected by edges. In a wide array of disciplines, data can be intuitively cast into this format [1]. For example, computer network consist of routers/computers (nodes) and the links (edges) between them. Social networks consist of particular biological function. Graphs are seemingly ubiquitous. The problems of detecting abnormalities (outliers) in a given graph and of generating synthetic but realistic graphs have received considerable attention recently[3].

Identifying tightly coupled pattern to the problem of finding the distinguishing characteristics of real-world graphs, that is, the patterns that show up frequently in such graphs and can thus be considered as marks of realism. A good generator will create graphs which match these patterns. Patterns and generators are important for many applications [4].Detection of abnormal sub graphs/edges/nodes. Abnormalities should deviate from the normal patterns so understanding the patterns of naturally occurring graphs is a prerequisite for detection of such outliers [6].Simulation studies. Algorithms meant for large real-world graphs can be tested on synthetic graphs which look like the original graphs. [5].Realism of samples. We might want to build a small sample graph that is similar to a given large graph. This smaller graph needs to match the patterns of the large graph to be realistic. Graph compression. Graph patterns represent regularities in the data. This can be used to better compress the data [7].

Manuscript received August 25, 2013.

M. Antony Sundarsingh, Research Scholar,M.S University Tirunelveli,Tamilnadu,India.

Dr.S.P.Victor,HOD-CS,St.Xaviers College,Tirunelveli Tamilnadu, India.

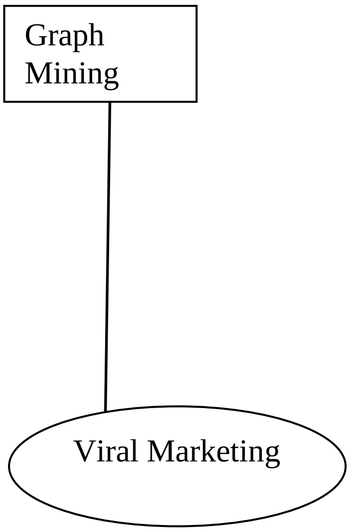

Fig 1: Application of Graph Mining

II.PROPOSED METHODOLOGY

Viral marketing is an application of social network mining that explores how individuals can influence the buying behavior of others. Traditionally, companies have employed direct marketing (where the decision to market to a particular individual is based solely on her characteristics) or mass marketing (where individuals are targeted based on the population segment to which they belong). These approaches, however, neglect the influence that customers can have on the purchasing decisions of others.

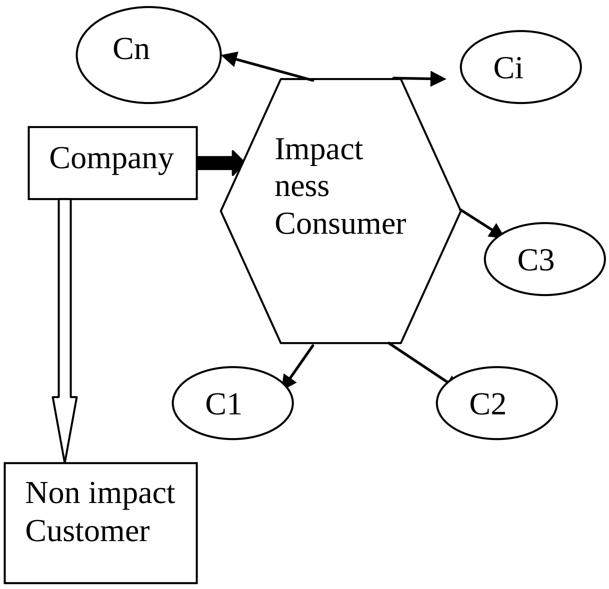

Fig 2: Viral marketing architecture for Graph mining

For example, consider a person who decides to use a particular mobile and persuades a group of friends to use the same mobile. Viral marketing aims to optimize the positive word-of-mouth effect among customers. It can choose to spend more money marketing to an individual if that person has many social connections. Thus, by considering the interactions between customers, viral marketing may obtain higher profits than traditional marketing, which ignores such interactions.

The growth of the Internet over the past two decades has led to the availability of many social networks that can be mined for the purposes of viral marketing. Examples include e-mail mailing lists, UseNet groups, on-line forums, instant relay chat (IRC), instant messaging, collaborative filtering systems, and knowledge-sharing sites. Knowledge sharing sites allow users to offer advice or rate products to help others, typically for free. Users can rate the usefulness or "trustworthiness" of a review, and may possibly rate other reviewers as well. In this way, a network of trust relationships between users (known as a "web of trust") evolves, representing a social network for mining. The network value of a customer is the expected increase in sales to *others* that results from marketing to that customer. In the example given, if our customer convinces others to use a certain mobile, then the mobile brand is justified in spending more money on promoting the brand to the particular customer. If, instead, our customer typically listens to others when deciding what mobile to use, then marketing spent on the customer may be a waste of resources. Viral marketing considers a customer's network value. Ideally, we would like to mine a customer's network (e.g., of friends and relatives) to predict how probable he/she is to buy a certain product based not only on the characteristics of the customer, but also on the influence of the customer's neighbors in the network. If we market to a particular set of customers then, through viral marketing, we may query the *expected profit from the entire network*, after the influence of those customers has propagated throughout. This would allow us to search for the optimal set of customers to which to market.

Considering the network value of customers (which is overlooked by traditional direct marketing), this may result in an improved marketing plan[8].

Given a set of n potential customers,

The neighbors of X_i are the customers who directly influence X_i. Then

X_i = 1 if customer i purchases the product being marketed,

0 otherwise.

M_i is defined as the *marketing action* that is taken for customer i. M_i may be continuous-valued (indicating the size of the discount offered, for example).

M_i= 1 if the customer is sent a coupon

0 otherwise

We would like to find the marketing plan that maximizes profit. A probabilistic model was proposed that optimizes M_i as a continuous value. That is, it optimizes the amount of marketing money spent on each customer, rather than just making a binary decision on whether to target the customer. The model considers the following factors that influence a customer's network value.

First, the customer should have high connectivity in the network and also give the product a good rating. If a highly-connected customer gives a negative review, her network value can be negative, in which case, marketing to her is not recommended.

Second, the customer should have more influence on others (preferably, much more) than they have on her.

Third, the *recursive* nature of this word-of-mouth type of influence should be considered. A customer may influence acquaintances, which in turn, may like the product and influence other people, and so on, until the whole network is reached.

The task of finding the optimal set of customers is formalized as a well-defined optimization problem:

Find the set of customers that maximizes the net profits. This problem is known to be NP-hard (intractable); however, it can be approximated within 63% of the optimal using a simple hill-climbing search procedure[8].

III. PROPOSED METHODOLOGY IMPLEMENTATION

The problem of graph matching is that of finding either an approximate or a one-to-one correspondence among the nodes of the two graphs. This correspondence is based on one or more of the following structural characteristics of the graph:

(1) The labels on the nodes in the two graphs must be same.

(2) The existence of edges between corresponding nodes in the two graphs *should* match each other.

(3) The labels on the edges in the two graphs should match each other.

These three characteristics may be used to define a matching between two graphs such that there is a one-to-one correspondence in the structures of the two graphs. In *exact graph matching*, we attempt to determine a one to-one correspondence between two graphs. Thus, if an edge exists between a pair of nodes in one graph, then that edge must also exist between the corresponding pair in the other graph. This may not be very practical in real applications in which *approximate matches may exist*, but an exact matching may not be feasible. Therefore, in many applications, it is possible to define an objective function which determines the similarity in the mapping between the two graphs. Fault tolerant mapping is a much more significant application in the graph domain, because common representations of graphs may have many missing nodes and edges. This problem is also referred to as *inexact graph matching*. Most variants of the graph matching problem are well known to be NP-hard. The most common method for graph matching is that of tree-based search techniques. In this technique, we start with a seed set of nodes which are matched, and iteratively expand the neighborhood defined by that set. Iterative expansion can be performed by adding nodes to the current node set, as long as no edge constraints are violated. If it turns out that the current node set cannot be expanded, then we initiate a backtracking procedure in which we undo the last set of matches. Let us consider few default network models for our proposed methodology with necessary assumptions [8].

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958, Volume-2, Issue-6, August 2013

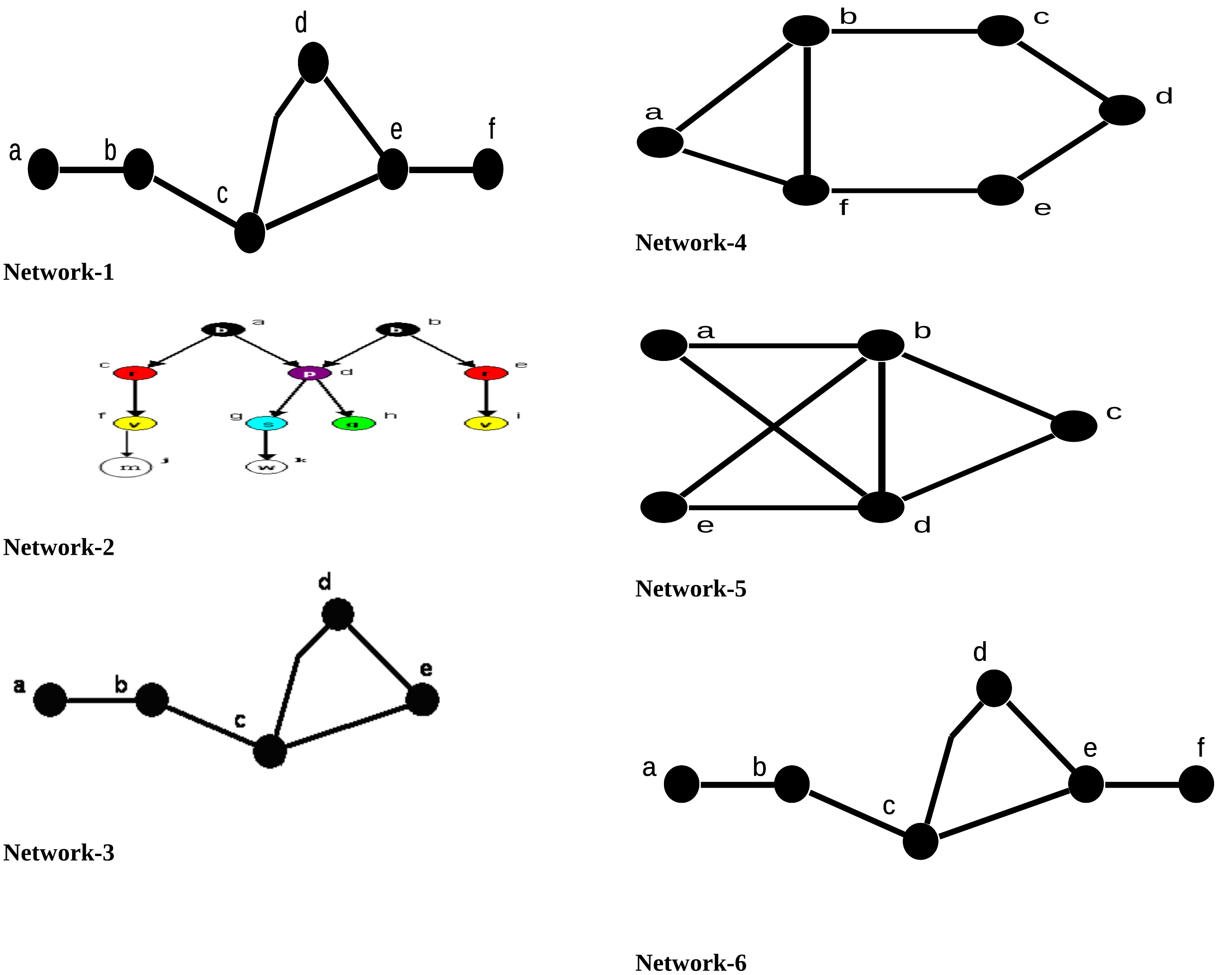

Fig 3: Micro Network models for viral computing

The proposed Optimization problem model is as follows,

Step 1: Collect the nodes from each network and set them as feasible as a default feature.

Step 2: Compute the Direct/Indirect influence factor for the feasible nodes.

Step 3: Identify the Feedback value for the node as positive or negative impact towards the network.

Step 4: Compute the impact value ranges from 0 to 1 based on their network weightage.

Step 5: Select the optimized node -Z

X = Max(Di/Dj) ∀ i, j ε N

Y = X ∩ {FB=1}

Z = Max (ImpVal I,Imp Val j) ∀ i, j ε N

Priority Network=Max (Z ∩ Inf (Di/Dj))

Step 6: Apply the Managerial action based on the Network priority.

IV. IMPLEMENTATION RESULT

Now applying the Graph matching pattern analysis we attain the following results, Figure1 and Figure 5 are Exact graph matches whereas

Figure 1 and Figure 3 are inexact graph matches with maximum level of approximation.

So eliminating the two duplicated networks we collect the remaining graph networks for identifying the exact network with appropriate priorities. The computations are as follows,

Table 1: Computation Table for the Proposed Model

Network ID	Feasible Nodes	X=Influence (Direct/Indirect)	Y=Feedback 1= + ve 0-= - ve	Impact value for PF	Z=Optimized Node	Expected Coverage	Preference Network Priority
Network 1	**a**	1 / 5	0	0.5	C	60 %	2
	b	2 / 5	1	0.4			
	c	**3 / 5**	**1**	**0.7**			
	d	2 / 5	0	0.3			
	e	**3 / 5**	0	0.4			

	f	1 / 5	1	0.5			
Network 2	**a**	**2 / 5**	0	0.2	B	50 %	3
	b	**2 / 4**	**1**	**0.4**			
	c	1 / 1	1	0.5			
	d	2 / 1	1	0.5			
	e	1 / 0	0	0.2			
	f	1 / 0	0	0.3			
	g	1 / 0	1	0.2			
	h	0 / 0	0	0.0			
	i	0 / 0	1	0.0			
	j	0 / 0	1	0.0			
	k	0 / 0	1	0.0			
Network 4	**a**	2 / 5	0	0.4	F	60 %	2
	b	**3 / 5**	**1**	0.5			
	c	2 / 5	0	0.4			
	d	2 / 5	1	0.3			
	e	2 / 5	1	0.2			
	f	**3 / 5**	**1**	**0.7**			
Network 5	**a**	2 / 4	1	0.2	D	100 %	1
	b	**4 / 4**	0	0.8			
	c	2 / 4	1	0.3			
	d	**4 / 4**	**1**	**0.7**			
	e	2 / 4	0	0.2			

Network 5 consumes more priority than with the network 1 and network 4 finally network 2 will be considered for the maximum coverage of business.

V.CONCLUSION

In this paper we describe the methodology for implementing a well defined problem to a Graph pattern, These are only some of the models with modular nature to implement it directly; there are many other models which add new ideas or combine existing models in novel ways. We have looked at many of these and discussed their strengths and weaknesses. The overall method proves to be highly efficient compared to mining significant and open trees, dramatically reducing running time and number of features mined. Moreover, the experimental results revealed that the expressiveness of Graph matching Node impact influence optimization representatives is significantly higher than that of open trees, because a lower number of features are associated with better accuracy, mainly due to higher specificity, reducing false alarms in classification tasks. In the future we will extend our research to propose a Graph-Analysis Implementer for any real-time complex entities.

REFERENCES

[1] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Ma-honey. Statistical properties of community structure in large social and information networks. In WWW, pages 695-704, 2008.

[2] C. Liu, F. Guo, and C. Faloutsos. Bbm: Bayesian browsing model from petabyte-scale data. In KDD, pages 537-546, 2009.

[3] Y. Low, J. Gonzalez, A. Kyrola, D. Bick son, C. Guestrin, and J. M. Heller stein. Graph lab: A new framework for parallel machine learning. In UAI, pages 340-349, 2010.

[4] R. Gemulla, E. Nijkamp, P. Haas, and Y. Sisma-nis. Large-scale matrix factorization with distributed stochastic gradient descent. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 69-77. ACM, 2011.

[5] A. Ghoting, R. Krishnamurthy, E. P. D. Pednault,B. Reinwald, V. Sindhwani, S. Tatikonda,Y. Tian,and S. Vaithyanathan. System: Declarative machine learning on map reduce. In ICDE, pages 231-242, 2011

[6] U. Kang, H. Tong, J. Sun, C.-Y. Lin and C. Faloutsos.Gbase: an ancient analysis platform for large graphs.VLDB J., 21(5):637-650, 2012.

[7] Dr.S.P.Victor,Antony Sundar singh:" Design and Development of Abstractness in Graph Mining Technique using Structural Datum "-IJSCE-Vol-3,Issue-3-Jun-2013.

[8] *http://www.elsevierdirect.com/companions/9780 123814791/chapters_from_the_second_edition/chapter_9.pdf*