Analytical Modeling of Social Network Growth Using Multilayer Network Projection

[1] Kambiz Behfar, [2] Ekaterina Turkina, [3] Patrick Cohendet, [4]Animesh Animesh

[1, 2, 3] Department of International Business, HEC Montreal, H3T 2A7, Canada

[4] Department of Information Systems, Faculty of Management, McGill University, H3A 1G5, Canada

Abstract - In this study, we present a new network model to address the question of how social networks such as groups in social websites, networks of social events, online video game groups evolve in time. A common feature of all these networks is that first, new users attach to an existing group (clique), and second, the decision of new network users is greatly influenced by already-joined users. In our model, new user joins the underlying network from a parallel network layer as the consequence of influence of underlying network users on the ones from other networks. This influence could be any type of relationship exterior to the underlying network layer such as friendship. For the new network model design, we utilize the concepts behind: 1. Growing network model presented by Barabási–Albert (BA), 2. Social influence-driven contagion model and 3. Multiplex network model. The way we treat this challenging phenomena is by multilayer projection, i.e. friends of each actual user that exist in other social network layers are projected as virtual users onto the underlying network layer. At each time step, each virtual user might become actual user based on a proposed Social influence-driven contagion method.

Keywords - *Social network, growth, contagion, Multilayer*

1. Introduction

Social networks grow fast through the addition of new interactions or new adoptions. Barabási–Albert (BA)model assumes time-homogeneous addition of user into network (one user, m links per time step) with preferential attachment mechanism. This results in a scale-free network degree evolution.

In addition, networks could evolve by new adoption using peer-influence-driven contagion model or homophily based diffusion models. At the same time, growth of social networks is not limited to just one network layer, since people in life situations are in different social networks, and those social networks are interdependent. Multiplex network models aim to explain growth of these interdependent networks. In the following, we will discuss all these concepts.

1.1 Network Evolution Models

There have been primary contributions in the area of network models, non-growing randomly connected network model presented by Erdős–Rényi (ER) [1], non-growing randomly re-connected network model (so-called small world) presented by Watts and Strogatz (WS) [2], and growing network with probability of addition of new user proportional to the number of incoming links (so-called preferential attachment model) presented by Barabási–Albert (BA) [3]. In ER and WS models, the number of users in network is fixed, and linkages among existing users are formed, while BA model assumes time-homogeneous addition of user into network with preferential attachment mechanism. Barabási–Albert have claimed that a common property of many large networks in that the vertex connectivity follows scale-free power-law distribution, and concluded that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems [1][4]. At the same time, preferential attachment does not always explain network evolution. For instance, in real systems a user's connectivity and growth rate does not depend just on its age. As an example, some web documents with good content and good marketing obtain much attention in a very short time. To address this concern, scientists have introduced fitness method(fit-get-richer instead of rich-get-richer) to explain the probability of user addition to network[5][6][7].

1.2 Adoption Models

Data mining helps in the process of direct marketing by producing prediction models of potential customer's response to the marketing based on past behavior or demographic information. However, data mining assumes that each customer makes decision independent of others; while in reality this is not the case. Each individual's decision on the purchase of a product is influenced by neighbours or acquaintances. Data mining plays important role in the target marketing process by constructing models to predict customer adoption response given characteristic information. Data mining assumes that all individuals make adoption decision separately, while viral marketing based on word of mouth takes into consideration network value of customers, i.e. individual's decision on purchase of a product is influenced by neighbours. In addition, viral marketing is a beneficial marketing strategy in terms of cost, because customers themselves take over promotional efforts.

Models of finding adoption probabilities and spread of adopting behavior have attracted much attention from both academia and industry such as in business [8], viral marketing [9] and demand prediction [10]. Typical word of mouth simulations use random graphs rather than real networks, and do not consider adoption propensities, correlation between individual characteristics, and similarities between connected neighbours. At the same time, recent adoption models consider influence-driven contagion and homophily-based diffusion. Influence-driven contagion implies that individual's probability of adoption at each time step depends on his/her neighbours' adoption at the previous time step [11]. Let v be a non-yet-adopter and U contain all's neighbours, then the probability that v adopts at time $t+1$ is given below; where $P_{u,v}$ represents the probability of u influencing$v$, assuming that all neighbours have equal probability of influencing, then $P_{u,v} = 1/k_u$ [11].

$$P_v = 1 - \prod_{u \epsilon U} (1 - P_{u,v}) = 1 - \prod_{u \epsilon U} (1 - 1/k_u) \tag{1}$$

Apart from influence-driven contagion, structural characteristic of network entity affects the probability of adoption such as structural equivalence (homogeneity of adopters) i.e. people with similar social ties or connecting to same individual or group show similar opinion of adoption[12].Aral et al. have defined a propensity measure which is a function of characteristic-based preference and peer influence. This originally comes from Bass diffusion model, in which new product adoption is governed by innovation rate as well as imitation rate [13][14][15].

1. 3 Multiplex Network Models

Social networks represent superposition of inter-dependent networks; where users denote individuals and links account for social ties. Interactions among these multiplex networks have been studied in few papers. Multiplex networks are structured in multilayers; where interconnections represent interactions between one user from one layer and its counterparts in another layer. In one paper, diffusion dynamics of multiplex networks has been studied using Laplacian matrix for each layer[16], in another paper, same authors have used game theory to model the evolution of cooperation in multiplex networks [17], in another paper, some others have discussed preferential attachment (BA) in the interaction between interdependent networks by proposing a class of stochastic models where interdependent users are connected via preferential attachment [18].

1.4 Social Network Growth Model

We seek to address the question of how social networks such as groups in social websites, network of social events, online video game groups grow in time, utilizing multilayer network projection modeled using exogenous as well as network intrinsic factors. We utilize the concepts behind 1. Growing network model presented by Barabási–Albert (BA), 2. Social influence-driven contagion model and 3. Multiplex network model. Table I reports the associated network models.

Table I: Different network models

	Evolution models	Adoption models	Multiplex models
Goal	Simulation of net evolution	Viral marketing, market targeting	Interdependent network evolution
Model I	Erdős–Rényi [1] Random network	Influence-driven contagion [15]	game theory of cooperation [17]
Model II	Watts &Strogatz, [2] small world	Homophily-based diffusion [14]	Diffusion dynamic , LaplacianM. [16]
Model III	Barabási-Albert[3] Pref. attachment	Propensity based method[13]	Epidemic and percolation theory[19]
Model IV	Barabási[5], Cald-arelli[6]fitness model,Caldarelli[7]Varying fitness	structural characteristic structural equivalence[12]	preferential attachment in interdependent networks [18]

Unlike BA network model investigating the attachment of new user to an existing user, our model explains the attachment of new user to a clique (connected group), such as group of friends or group of online gamers. Therefore, this model does not hold for individual-based networks such as citation or co-authorship networks.

Unlike peer influence-driven contagion which is based on user influencing or causing outcome on its neighbour, or homophily-based diffusion where similarities between users create correlated outcome patterns among neighbours, in our proposed network model, probability of new user joining the underlying network depends on prior connections with existing groups in the underlying network.

Unlike multiplex network, where its diffusion dynamics is based on mutual interactions among neighbouring network layers (using Laplacian matrix where its diagonal elements represent intra-layer interactions and its off-diagonal elements represent inter-layer interactions), in this new network model there is one underlying network layer, and users from other layers are projected on this underlying layer, so interactions are not mutual between layers.

In the proposed network model, a new user comes from another network layer with prior relationship with the underlying network layer. The underlying network layer could be an online-video-game and other layers could be the actual game subscribers' friends on any social websites. An individual joins a group of gamers if he/she has prior links with a few gamers within the game group. In the case of a public event in social website, individuals join the event if they know a couple of friends already attend the event. A common feature of these examples is that first, a new user joins a clique (connected group), and second, a new user situates in another network layer (e.g. social network) with prior relationship to the underlying network layer (e.g. online group of gamers). The probability of new user joining to the network is obtained by a proposed social influence contagion model.

2. Model design

Think of a few practical examples such as case of Facebook event invitation where we attend an event, not necessarily because this event is desirable to us, but because our friends attend the event as well. We are in fact common friends of those event attendees or in their friend list, which is a network layer parallel to the event network layer. In the case of online video games, we play an online game not only because the game is amazing, but because our friends play the group game as well. A common feature in all these examples is that the new user lies in a network layer, which has prior connection with the underlying network layer. This connection could be through mutual friends. Therefore, this new network model has some exclusive features:

1) This network model is based on existing groups of users (clique) in its very essence, as seen in Figure 1.

2) Network properties depend on both internal degree of each user (k_i^j user j group i) and external degree (k'^j_i) which represent the number of links to other networks. However, it is independent of the properties of these other network links.
3) Users are divided into actual and virtual users; where actual users belong to the underlying network layer, and virtual users are the ones projected onto the underlying network layer. As shown in Figure 2, we have network layers L_i (with actual users) and L_{i+1} (with virtual users), and two layers possess few users in common, which are projected onto the underlying layer L_{i+2}.

2.1 Network Layer Projection

The analysis of this new network model is challenging, because each network user has many unknown friends NOT in the underlying network layer, BUT in other social network layers. For instance, each online video gamer has most of his/her friends in other social networks who have not yet adopted this online game. As soon as a few of his/her friends adopt the game, this influences his/her adoption behaviour. In another case (as seen in Figure 1), if few groups of individuals attend a public event, each attendee has a group of external friends. As long as few of their mutual friends attend the event, this affects their decision to attend too.

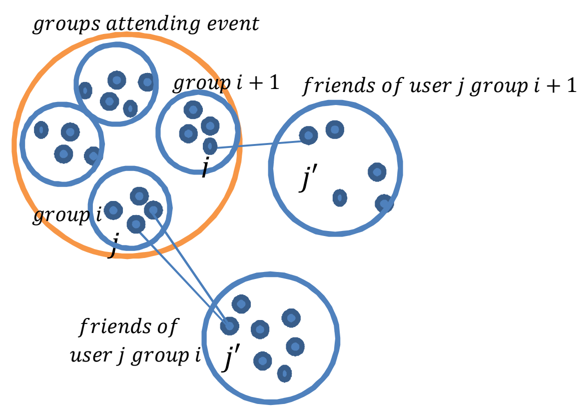

Figure 1: Friends of user ***j*** from group ***i*** of underlying network situate in other network layers.

Therefore, we treat this challenging phenomena by multilayer projection, i.e. friends of each user lying on other networks are projected as virtual users onto the underlying network layer containing groups of actual users.

At each time step, each virtual user with more than certain number of links to actual group of users (users) becomes an actual user using influence-based contagion method. At next time step, the actual users update their numbers, and procedure follows similarly. User j belonging to group i is represented by g_i^j with actual degree of k_i^j, and virtual degree of k'^j_i . If it is a virtual user, it is denoted by $g_i^{j'}$, its actual degree is denoted by $k_i^{j'}$, and its virtual degree is denoted by $k'^{j'}_i$. The number of users in group i is represented by n_i , and after t time steps, it is given by $n_i(t)$.

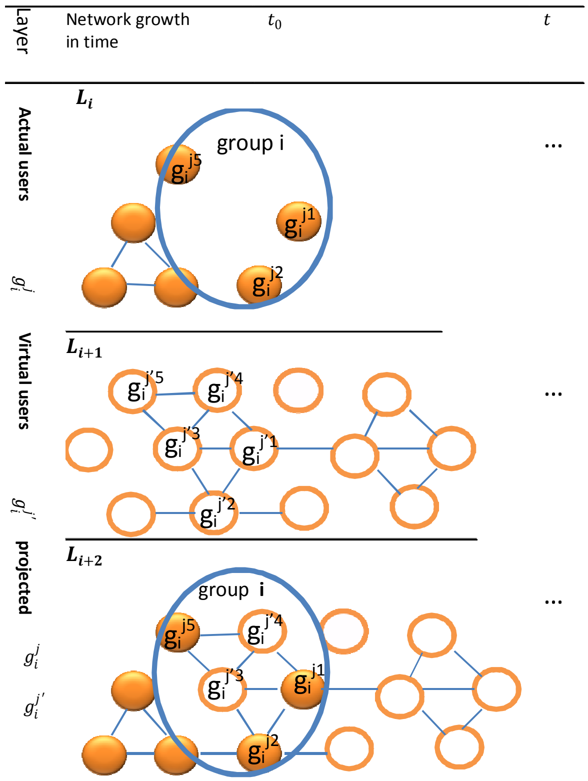

Figure 2: Network layers L_i(with actual users) and L_{i+1}(with virtual users), projected onto layerL_{i+2}(actual and virtual)

2.2 Growth Model

Network growth is categorized into topological growth and size growth, where we discuss network growth in size by adding new users to the underlying network using a proposed influence-based contagion method.

2.2.1 Influence-Based Contagion Method

We describe how network growth can be formulated using a proposed influence-based contagion method, in a sense that if the actual degree of user j', denoted by $k_i^{j'}$,is greater than the ratio of virtual degree of user j', denoted by $k'^{j'}_i$, this means that the fraction of total user j'neighbors have already joined the underlying network; this causes user $j'$to also join the underlying network, indicated by

$$\delta\left(\frac{k_i^{j'}}{k'^{j'}_i}-c\right)=\begin{cases}1 & if\ \ k_i^{j'}>c.k'^{j'}_i\\ 0 & otherwise\end{cases} \qquad (2)$$

where c is a constant. The Actual degree k_i^j for user g_i^jevolves as below:

$$k_i^{1j}=k_i^{0j}+\sum_{j'=1}^{j'=k'^j_i}\delta\left(\frac{k_i^{j'}}{k'^{j'}_i}-c\right) \qquad (\Delta t=t_1-t_0)$$

$$k_i^{2j}=k_i^{0j}+\Delta k_i^{1j}+\sum_{j'=1}^{j'=k'^j_i-\Delta k_i^{1j}}\delta\left(\frac{k_i^{j'}}{k'^{j'}_i}-c\right) \qquad (\Delta t=t_2-t_1)$$

$$k_i^{3j}=k_i^{0j}+\Delta k_i^{1j}+\Delta k_i^{2j}+\sum_{j'=1}^{j'=k'^j_i-\Delta k_i^{1j}-\Delta k_i^{2j}}\delta\left(\frac{k_i^{j'}}{k'^{j'}_i}-c\right) \qquad (\Delta t=t_3-t_2)$$

$$\vdots$$

$$k_i^{(n+1)j}=k_i^{0j}+\Delta k_i^{1j}+\cdots+\Delta k_i^{nj}+\sum_{j'=1}^{j'=k'^j_i-\Delta k_i^{1j}+\cdots+\Delta k_i^{nj}}\delta\left(\frac{k_i^{j'}}{k'^{j'}_i}-c\right) \qquad (\Delta t=t_{n+1}-t_n) \qquad (3)$$

Considering the approximate mean of $\langle\Delta k_i^j\rangle_{t_1}^{t_n}$, equation (3) can be re-formulated as below:

$$k_i^{(n+1)j}=k_i^{0j}+n\,\langle\Delta k_i^j\rangle_{t_1}^{t_n}+\sum_{j=1}^{j'=k'^j_i-n\langle\Delta k_i^j\rangle_{t_1}^{t_n}}\delta\left(\frac{k_i^{j'}}{k'^{j'}_i}-c\right) \qquad (4)$$

To obtain$k_i^{(n+1)j}$, we require a formula which depends on just initial values of parameters.$\langle\Delta k_i^j\rangle_{t_1}^{t_n}$can be obtained approximately by averaging over few time steps. However, we need to approximate the sum$\Sigma_{j'=1}^{j'=k'^j_i-n\langle\Delta k_i^j\rangle_{t_1}^{t_n}}\delta\left(k_i^{j'}/k'^{j'}_i-c\right)$ as well. We see the modified contagion method in the next subsection.

2.2.2 Modified Contagion Method

As compared to the previous section, here we substitute$p_i^j k'^j_i$ for $\Sigma_{j'=1}^{j'=k'^j_i}\delta\left(k_i^{j'}/k'^{j'}_i-c\right)$, where p_i^j is the probability of joining the fraction of user $j$virtual neighbors to the underlying network and becoming actual users.

$$k_i^{1j}=k_i^{0j}+p_i^j k'^j_i \qquad (\Delta t=t_1-t_0)$$

$$k_i^{2j}=k_i^{0j}+\Delta k_i^{1j}+p_i^j\left(k'^j_i-\Delta k_i^{1j}\right) \qquad (\Delta t=t_2-t_1)$$

$$k_i^{3j}=k_i^{0j}+\Delta k_i^{1j}+\Delta k_i^{2j}+p_i^j\left(k'^j_i-\Delta k_i^{1j}-\Delta k_i^{2j}\right) \qquad (\Delta t=t_3-t_2)$$

$$\vdots$$

$$k_i^{(n+1)j} = k_i^{0j} + \Delta k_i^{1j} + \cdots + \Delta k_i^{nj} + p_i^j \left(k'^j_i - \Delta k_i^{1j} - \cdots - \Delta k_i^{nj} \right) \quad (\Delta t = t_{n+1} - t_n) \qquad (5)$$

Considering the approximate mean of$\langle \Delta k_i^j \rangle$, equation (5) can be re-formulated as below:

$$k_i^j(t) = k_i^j(0) + t\langle \Delta k_i^j \rangle + p_i^j \left(k'^j_i(0) - t\,\langle \Delta k_i^j \rangle \right) \qquad (6)$$

This$k_i^j(t)$ depends on just initial values of parameters. Total number of users in group i at time t, $n_i(\mathrm{t})$ is obtained as:

$$n_i(\mathrm{t}) = n_i(0) + \sum_{j=1}^{j=k_i^j(t)} t\langle \Delta k_i^j \rangle \qquad (7)$$

In next subsection, we discuss how to calculate p_i^j.

2.2.3 Probability p_i^j Determination

Here we discuss possible methods to obtain p_i^j, as to substitute in (5). Comparing (3) and (5), we obtain:

$$\sum_{j'=1}^{j'=k'^j_i} \delta \left(\frac{k_i^{j'}}{k'^{j'}_i} - c \right) = p_i^j \left(virtual\ node\ j' \rightarrow actual\ node\ j \right) k'^j_i \qquad (8)$$

In the first approach, consider the proposed influence-based contagion model, the probability of j'joining the underlying network should be proportional to the fraction$k_i^{j'}/k'^{j'}_i$. One can tune this proportionality with a parameter μ_i^jas shown in(9).

$$1.\ p_i^j = \mu_i^j \sum_{j'=1}^{j'=k'^j_i} \frac{k_i^{j'}}{k'^{j'}_i} \qquad (9)$$

In the second approach, the probability of user j'joining the underlying network is$1 - \prod_{j' \in k_i^j} (1 - 1/k_i^{j'})$obtained by (1). Therefore, p_i^jshould be proportional to this probability. One can tune this proportionality with a parameter μ_i^jas shown in(10).

$$2.\ p_i^j = \mu_i^j \sum_{j=1}^{j'=k'^j_i} \left[1 - \prod_{j' \in k_i^j} \left(1 - \frac{1}{k_i^{j'}} \right) \right] \qquad (10)$$

3. Empirical Applications

Groups in social websites, network of social events, online video game groups are just few examples of empirical applications of this new network model. A common feature of all these networks is that first, new users attach to an existing group (clique), and second, the decision of new users joining to the underlying network is greatly influenced by already-joined users. See few applications listed below:

3.1 Viral Marketing in Social Network

A new social (group-based) website is launched. It represents a social network with initial few subscribers. To attract subscribers, the firm launching the website does not have any choice other than to endure huge diverse advertisements. Using this new network model, and producing a map of how the network of subscribers grows in time, firms can do direct marketing instead of mass marketing. This model can be operationalized using Search engines looking for friends of subscribers in other social websites such as facebook, linkedin, …(Wajam[20])

3.2 Marketing map for Game Publishers

Online game publisher launches a new online game. There are initially few gamers, but the game producer has made a huge bet to have its product spread worldwide. At this moment, the publisher's best shot is to partner with popular social websites, TV, marketing and monetization firms [21] to do advertisement while enduring additional huge marketing cost. On the other hand, customers, mostly the youth, have a lot of free-to-play choices. Why would they bother to subscribe unless it is a game worth paying for? Ournew network model can produce a map of how network of subscribers grows, leading to an optimal marketing strategy. Then, firms can do direct marketing instead of mass marketing.

4. Conclusion

We presented a new network model to address the question of how social networks such as groups in social websites, network of social events, online video game groups grow in time. In this network model, a new user attaches to a network layer from another network layer as the consequence of influence of underlying network users on the ones from other networks. We treated this challenging phenomena by multilayer projection, i.e. friends of each user belonging to other networks were projected as virtual users onto the underlying network layer containing the groups of actual users.
At each time step, each virtual user with more than certain number of links to the actual group of users becomes an actual user using the proposed influence-based contagion method, and in the modified contagion method, we substituted$p_i^j k'^j_i$for $\sum_{j'=1}^{j'=k'^j_i} \delta \left(k_i^{j'}/k'^{j'}_i - c \right)$, where p_i^j is the

probability of fraction of user j virtual neighbours joining the underlying network and becoming actual users.

References

[1] Erdős, P. andRényi,A. (1960) "On the evolution of random graphs" Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5: 17–61

[2] Watts, D. J.; Strogatz, S. H. (1998) "Collective dynamics of 'small-world' networks" Nature 393 (6684).

[3] Barabasi, A.L. and Albert,R. (1999) “Emergence of scaling in random networks” Science Vol 286

[4] Barabasi, A.L., Alberta, R., and Jeong,H. (2008) “Mean-field theory for scale-free random networks” Physica A, 272

[5] Bianconi, G. andBarabasi, A. L.(2001) “Competition and multiscaling in evolving networks” EurophysLett. 54, 436

[6] Servedio,V. D. P. and Caldarelli,G.(2004) “Vertex Intrinsic Fitness: How to Produce Arbitrary Scale-Free Networks” Phys Rev E 70, 056126

[7] Caldarelli,G.,Capocci, A., Rios,P. D. L., and Munoz,M. A.(2002) “Scale-free networks from varying vertex intrinsic fitness.” Phys. Rev. Lett. 89, 258702

[8] Liben-Nowell,D.,Kleinberg,J. (2008) “Tracing Information Flow on a Global Scale Using Internet Chain-Letter Data. Proc. National Academy of Sciences, 105(12):4633–4638

[9] Domingos, P. and Richardson, M. (2001) "Mining the Network Value of Customers," Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining (pp. 57-66), San Francisco, CA: ACM Press.

[10] Hartmann, W.R., (2010) "Demand Estimation with Social Interactions and the Implications for Targeted Marketing," Marketing Science, 29(4).

[11] Chen W, Wang Y, Yang S.(2009)“Efficient influence maximization in social networks” Proc. 15th ACM SIGKDD Internat. Conf. Knowledge Discovery and Data Mining

[12] Fang,X. (2013) “Predicting Adoption Probabilities in Social Networks” Information systems Research, Vol. 24

[13] Aral, S., Muchnik, L. and Sundararajan, A. (2013) “Engineering Social Contagions: Optimal Network Seeding in the Presence of Homophily” Network Science.

[14] Aral,A. and Walker, D. (2011) “Creating Social contagion through viral product design: A randomized trial of peer influence in networks” Management Science, 57(9), Sep 1623-1639.

[15] Aral,A. and Walker, D. (2012) “Identifying influential and susceptible members of social networks” Science, 337 (6092): 337-341

[16] Gomex, S. et al (2013) “Diffusion dynamics on multiplex networks” Phys Rev. Lett. 110, 028701

[17] Gomez-Gardenes, J. et al (2012) “Evolution of cooperation in multiplex networks”

[18] Podobnik, B. et al (2012) “Preferential attachment in the interaction between dynamically generated interdependent networks” A Letters Journal exploring the frontiers of physics (EPL) , 100, 50004.

[19] Son, S.W.,Bizhani,G., Christensen, C., Grassberger, P. , and Paczuski, M. (2012) “Percolation theory on interdependent networks based on epidemic spreading” A letter journal exploiting the frontiers of physics.

[20] Wajam:http://www.wajam.com/

[21] SuperRewards: http://www.superrewards.com/