Research Article

Yilun Shang*

Lie algebraic discussion for affinity based information diffusion in social networks

https://doi.org/10.1515/phys-2017-0083
Received May 17, 2017; accepted Sep 24, 2017

Abstract: In this paper we develop a dynamical information diffusion model which features the affinity of people with information disseminated in social networks. Four types of agents, i.e., susceptible, informed, known, and refractory ones, are involved in the system, and the affinity mechanism composing of an affinity threshold which represents the fitness of information to be propagated is incorporated. The model can be generally described by a time-inhomogeneous Markov chain, which is governed by its master (Kolmogorov) equation. Based on the Wei-Norman method, we derive analytical solutions of the model by constructing a low-dimensional Lie algebra. Numerical examples are provided to illustrate the obtained theoretical results. This study provides useful insights into the closed-form solutions of complex social dynamics models through the Lie algebra method.

Keywords: social network, Lie algebra, Markov process, diffusion

PACS: 02.20.Sv, 02.50.Ga, 87.23.Cc

1 Introduction

There has been recently a growing interest in human-activated spreading phenomenon or social dynamics partly due to the ever-increasing availability of large databases and the fast development of computational devices [1–4]. A key issue underlying such systems is to understand how the interactions between social agents create order in their behavior or state from an initially disordered one. In the early stage, population models based on master equations were borrowed from mathematical epidemiology [5] to describe the spreading of information; see e.g. [6–8]. A drawback of such models is that the ad hoc features of information diffusion in complex social networks are technically overlooked. This is partially alleviated by introducing some realistic aspects including human activities [9–11] and individual choices [12, 13].

Recently, some variant epidemiological models capturing the willingness of people to disseminate information have been reported. In [14], a SIR-like model (called message affinity model) was introduced by assigning to the substrate network nodes a propensity value representing their affinity with the message/news being disseminated. Here, S refers to the susceptible agents who have not received the information, I the informed agents who are propagating the information, and R the refractory agents who do not spread the information anymore. A susceptible agent touched by the message becomes informed only if her affinity value is higher than a certain threshold. Another variant called UKAE model was proposed in [15], where agents are divided into four groups: unknown (or susceptible), known, approved (or informed), and exhausted (or refractory). The newly added “known” state refers to the situation where the agent is aware of the message but not willing to transmit it since she is suspicious of the authenticity of the message. The model encodes memory effects and social reinforcement in that an individual judges whether the message is true (and forwards it to her neighbors) depending on the number of times she has heard it. An analogous four-compartment model (SKIR model) was examined in [16], where a known individual confirms the facticity of the message based on her self-judgement rather than the number of times she heard.

While a wealth of models have been developed to understand various social dynamics, most of them are based on either numerical simulations or empirical data, with the noteworthy exception of the Deffuant opinion interaction model analytically explored by several mathematicians [17–20]. However, analytical treatment is of interest in social science, mainly because real data are not always accessible and numerical simulations tend to be difficult for large systems. The objective of this paper is to derive analytical solution for an affinity based information diffusion model [16] with time-varying transmission rates by employing the Lie algebraic approach. This is achieved by first formulating the process as a time-inhomogeneous
Markov chain, whose master (or Kolmogorov) equation governs the evolution of the probability of the process being in a given state at a given time. We then construct a Lie algebra with a finite dimension which solves the master equation readily under the Wei-Norman’s framework [21, 22].

The rest of the paper is organized as follows. In Section 2, we describe the affinity based information diffusion model. The Wei-Norman method is briefly reviewed in Section 3. Section 4 shows the application of the Lie algebraic discussion to the model. Concluding remarks and some open problems are provided in Section 5.

2 Affinity based diffusion model

We consider a continuous-time information diffusion model, where each agent adopts one of the following four states: (i) Susceptible: the agent has not received the information; (ii) Known: the agent is aware of the information but not willing to transmit it since she doubts the authenticity of the information; (iii) Informed: the agent confirms the information and transmits it to all her neighbors; (iv) Refractory: after spreading the information, the agent loses interest and never transmits it again. The total population is denoted by \( N \). Let the numbers of agents at time \( t \) in each of the four states be \( S(t) \), \( K(t) \), \( I(t) \), and \( R(t) \), respectively. Thus, we have \( N = S(t) + K(t) + I(t) + R(t) \) for all \( t \geq 0 \).

As in [14, 16], we assign the affinity value \( a_i \in [0, 1] \) to each agent \( i \), which represents her propensity to engage in forwarding the information in question. \( \{a_i\}_{i=1}^N \) are independently identically distributed according to the uniform distribution. The substrate social network is regarded as an evolving network with average degree \( k(t) \) at time \( t \). The affinity threshold \( \theta(t) \in [0, 1] \) is interpreted as the information fitness to trigger the activation—an individual \( i \) can change her state from susceptible to informed or known only if \( a_i \geq \theta(t) \). This is of course reminiscent of the confidence threshold studied in the Deffuant model [18, 19], but the meaning here is totally different.

The information diffusion with respect to continuous-time \( t \) in the network is governed by the following rules: (i) A susceptible agent will interact with an informed agent in her neighborhood at the time points of a unit Poisson process, when her state becomes informed with probability \( 1 - p_1(t) \) and known with probability \( p_1(t) \), respectively, if her affinity is higher or equal to \( \theta(t) \). Otherwise, the susceptible agent will become refractory immediately. (ii) A known agent will interact with an informed agent in her neighborhood at the time points of a unit Poisson process, when she becomes informed with probability \( p_1(t) \) and remains known with probability \( 1 - p_2(t) \). (iii) An informed agent will become refractory with probability \( p_2(t) \) and remains informed with probability \( 1 - p_3(t) \) at the time points of a unit Poisson process. (iv) Refractory agents stay unchanged. The transition diagram is illustrated in Figure 1. The diffusion mechanism is slightly different from the SKIR model studied in [16] (by putting \( p_3 \) therein, the transition probability from known to informed spontaneously, zero). The similar mean-field final-size equation [16, Eq. (6)] can be reproduced through the Wei-Norman method (see Eq. (21) below).

3 The Wei-Norman method

Some basic concepts of Lie algebra and the Wei-Norman method [21, 22] for solving time-inhomogeneous Markov chains are reviewed in this section for the completeness and reader’s convenience.

A Lie algebra is a vector space \( \mathcal{L} \) over some field \( F \) together with a bilinear map \( [,] : \mathcal{L} \times \mathcal{L} \to \mathcal{L} \), called the Lie bracket, satisfying \( [A,A] = 0 \) and the Jacobi identity

\[
[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0,
\]

for all \( A, B, \) and \( C \in \mathcal{L} \). For \( A \in \mathcal{L} \), an adjoint operator \( \text{ad}A \) is defined by

\[
(\text{ad}A)B = [A, B], \quad B \in \mathcal{L}.
\]

Hence, \( (\text{ad}A)^2 B = [A, [A, B]] \). Note that every associate algebra gives rise to a Lie algebra \( \mathcal{L} \) by defining the Lie bracket as a commutator

\[
[A, B] = AB - BA,
\]

where \( A, B \in \mathcal{L} \). In what follows, we will focus on this special Lie bracket. Employing the adjoint operator, the classical Baker-Campbell-Hausdorff formula can be written as.
where the matrix exponential is given by 
\[ e^A B e^{-A} = (e^{adA})B, \]
(4)

where \( A^t \) is the transpose of \( A \).

For a continuous-time Markov chain, taking values in a finite or countably infinite state space \( S \), its behavior is defined by a matrix \( Q(t) = (q_{ij}(t), i, j \in S) \), where \( q_{ij}(t) \) is the rate of transition from state \( i \) to state \( j \), for \( j \neq i \), and \( -q_{ii}(t) = q_i(t) = \sum_{j=1}^m q_{ij}(t) \) is the total rate at which the process moves out of state \( i \) at time \( t \). Using the Kolmogorov forward equation, the probability distribution of the process at time \( t \), \( p(t) = (p_i(t), i \in S) \), is given by

\[
\frac{dp(t)}{dt} = H(t)p(t),
\]
(5)

where \( H(t) = Q(t)^T \), and \( p(t) \) is a column probability vector with component \( p_i(t) \) describing the probability of finding the system in state \( i \) at time \( t \). Using the ‘ket’ notation \( |i\rangle \) [23, 24], the probability vector can alternatively be recast as

\[
|p(t)\rangle = \sum_{i \in S} P(i(t) | i) |i\rangle,
\]
(6)

where \( P(i(t) | i) \) is the probability that the Markov chain under consideration taking the value of \( i \) at time \( t \), and \( |i\rangle \) is a basis vector, linearly independent of any other basis vector with different value. The time-dependency of \( H(t) \) in (5) implies that the process is time inhomogeneous.

The Wei-Norman method requires a decomposition of the operator \( H(t) \) as

\[
H(t) = \sum_{i=1}^m a_i(t)H_i.
\]
(7)

where \( a_i(t) \) are real-valued functions, and \( H_i \) are linearly independent constant operators generating a Lie algebra \( \mathcal{L} = \text{span}\{H_1, \cdots , H_m\} \) by implementing a Lie bracket

\[
[H_i, H_j] = H_iH_j - H_jH_i = \sum_{k=1}^m \xi_{ij}^k H_k
\]
(8)

for some real \( \xi_{ij}^k \). It was shown that the solution of system (5) can be uncoupled into a product of exponentials:

\[
p(t) = e^{g_i(t)H_i} \cdots e^{g_m(t)H_m} P(0) := U(t)p(0),
\]
(9)

where \( g_i(t) \) are real-valued functions and \( g_i(0) = 0 \) for all \( i = 1, 2, \cdots , m \).

Substituting (7) and (9) into (5), we obtain

\[
\frac{dp(t)}{dt} = \sum_{i=1}^m a_i(t)H_iU(t)p(0)
\]
(10)

\[
= \sum_{i=1}^m \dot{g}_i(t) \left( \prod_{j=1}^{i-1} e^{g_j(t)H_j} \right) H_i \left( \prod_{j=i}^m e^{g_j(t)H_j} \right) p(0).
\]

Performing a post-multiplication by the inverse operator \( U^{-1} \) and repeatedly applying the Baker-Campbell-Hausdorff formula, we arrive at

\[
\sum_{i=1}^m a_i(t)H_i = \sum_{i=1}^m \dot{g}_i(t) \left( \prod_{j=1}^{i-1} e^{g_j(t)H_j} \right) H_i \left( \prod_{j=i}^m e^{g_j(t)H_j} \right) p(0),
\]
(11)

since the equation (10) holds for any \( p(0) \). Because the operators \( H_i \) are chosen to be linearly independent, we can compare the coefficients of each \( H_i \) in both sides of (11) to derive a set of ordinary differential equations for \( g_i(t) \) with initial values \( g_i(0) = 0 \) (involving \( \xi_{ij}^k \)).

The computational efficiency is a remarkable advantage of the Wei-Norman method since \( p(t) \) can be calculated in \( O(1) \) operations through (11) rather than \( O(t) \) by means of incremental direct integrations. Moreover, the matrix exponential form (9) is useful if the derivative of the solution with respect to a model parameter is needed [25].

4 Application of the Wei-Norman method

Through first principles described in Section 2, the probability vector for the information diffusion model can be written as

\[
|p(t)\rangle = \sum_{S,K,I} P(S, K, I | N, t) |S, K, I\rangle,
\]
(12)

where \( P(S, K, I | N, t) \) denotes the probability that from a cohort of size \( N \) there are \( S \) susceptible agents, \( K \) known agents, and \( I \) informed agents at time \( t \), leaving \( N-S-K-I \) refractory ones. \( |S, K, I\rangle \) is a basis vector, linearly independent of other basis vectors with different susceptible, known, and informed counts. The state space \( S \) is a finite set.

Let \( \hat{X} \) with a hat on it represent an endomorphism of the vector space spanned by the above basis vectors. The Kolmogorov equation governing the information diffusion process can be written as

\[
\frac{dp(t)}{dt} = H(t)p(t),
\]
(13)

with the time-evolution operator

\[
H(t) = k(t)(1 - \theta(t))(1 - p_1(t))\hat{\tau} \hat{S} + k(t)\theta(t)\hat{\tau} \hat{S} + k(t)(1 - \theta(t))p_1(t)\hat{\sigma} \hat{S} + k(t)p_3(t)\hat{\tau} \hat{K} + p_2(t)\hat{\delta} \hat{I},
\]
(14)

where \( \hat{\tau} |S, K, I\rangle = S|S - 1, K, I + 1\rangle. \]
Now we introduce another operator $\hat{\eta}$, defined as

$$\hat{\eta}|S, K, I\rangle = K|S, K, I\rangle,$$

so that the Lie algebra $\mathcal{L} = \text{span}\{\hat{\tau}, \hat{S}, \hat{\sigma}, \hat{I}, \hat{\rho}, \hat{\gamma}, \hat{\delta}, \hat{\eta}\}$ is closed under the action of the Lie bracket. We display in Table 1 the complete set of Lie brackets. The nine operators defined above are all linear operators acting on basis vectors. They are interpreted as follows: $\hat{\tau}$ returns the number of susceptible agents, depletes the susceptible population by one, and increases the known population by one; $\hat{S}$ returns the number of susceptible agents; $\hat{\sigma}$ returns the number of susceptible agents, depletes the susceptible population by one, and increases the known population by one; $\hat{I}$ returns the number of informed agents; $\hat{\rho}$ returns the number of known agents; $\hat{\gamma}$ returns the number of susceptible agents and depletes these by one; $\hat{\delta}$ returns the number of informed agents and depletes these by one; $\hat{\eta}$ returns the number of known agents and depletes these by one.

The remaining task is to apply the Wei-Norman method to Eq. (13). Namely, we want to look for a solution of the form

$$|p(t)\rangle = e^{g_1(t)\hat{\eta} + g_2(t)\hat{S} + g_3(t)\hat{K}} e^{g_4(t)\hat{\sigma}} e^{g_5(t)\hat{I}} e^{g_6(t)\hat{\rho}} e^{g_7(t)\hat{\gamma}} e^{g_8(t)\hat{\delta}} e^{g_9(t)\hat{\eta}}|p(0)\rangle.$$

Using (11) and the action of exponential operators shown in Table 2, we obtain the following linear relation

$$k(t)(1 - \theta(t))(1 - p_1(t))\hat{\tau} + k(t)\theta(t)\hat{\gamma} + k(t)(1 - \theta(t))p_1(t)\hat{\sigma} + k(t)p_3(t)\hat{\rho} - k(t)p_3(t)\hat{K} + p_3(t)\hat{\delta} - p_2(t)\hat{I} - k(t)\hat{S} = g_1(t)\hat{\eta} + g_2(t)\hat{S} + g_3(t)(\hat{K} + g_1(t)\hat{\eta}) + g_4(t)\hat{I} + \hat{g}_5(t)e^{g_4(t)\hat{K}}\hat{\tau} + \hat{g}_6(t)e^{g_4(t)\hat{\eta}}\hat{\sigma} + \hat{g}_7(t)e^{g_4(t)\hat{\rho}}\hat{I} + \hat{g}_8(t)e^{g_4(t)\hat{\gamma}}\hat{\delta} + \hat{g}_9(t)e^{g_4(t)\hat{\eta}}\hat{\eta}.$$
Assume that \( \langle \delta (t) \rangle = \Theta (t) \), and \( \Sigma \) is the density of susceptible agents in the population at time \( t \). Since \( R (\infty) \) (or \( s (\infty) ) \) is independent of \( p_1 \) and \( p_3 \) [16], we take \( p_1 = p_3 = 0 \) in (19), (20), and let \( t \) tends to infinity, which gives \( p (\infty) \) = \( e^{-k(1-\theta)/p_1} e^\delta p (0) \).

Finally, we derive the following equation:

\[ g (t) = \int _0 ^t e^{-\Gamma (u)} \left[ k(u) \varTheta (u) \right. \]

\[ - k(u)(1 - \varTheta (u))p_1(u)g_1(u) - p_2(u)g_3(u) \] du

\[ + \int _0 ^t p_2(u) e^{\Lambda (u)} (g_5(u) - g_6(u)g_7(u)) du, \]

where \( \Gamma (t) := \int _0 ^t k(u) du \), \( \Psi (t) := \int _0 ^t k(u)p_3(u) du \), and \( \Lambda (t) := \int _0 ^t p_3(u) du \).

In order to recover the final-size equation [16, Eq. (6)] for the SKIR model, we take \( p (0) = |N - 1, 0, 1 \) and assume that \( k(t) = k \), \( \varTheta (t) = \Theta \), and \( p_i(t) \equiv p_i \) for \( i = 1, 2, 3 \). Set \( S (t) = \sum _{S, K, I} s (S, K, I) \). Thus, \( s (t) := s (t)/N = \langle S (t)p (t) /N^2 \rangle \) is the density of susceptible agents in the population at time \( t \). Since \( R (\infty) \) (or \( s (\infty) ) \) is independent of \( p_1 \) and \( p_3 \) [16], we take \( p_1 = p_3 = 0 \) in (19), (20), and let \( t \) tends to infinity, which gives \( p (\infty) \) = \( e^{-k(1-\theta)/p_1} e^\delta p (0) \).

Finally, we obtain

\[ s (\infty) = \langle S (\infty)p (\infty)/N^2 \rangle = e^{-k(1-\theta)/p_1} (1 - s (\infty)). \]  

Note that Eq. (21) is essentially in line with Eq. (6) in [16] as desired.

Finally, we perform simulations on an Erdős-Rényi network and a small-world network with network size \( N = 5000 \) and average degree \( k = 10 \), respectively. For each data point, 1000 independent dynamical realizations are used to calculate the pertinent average values, which are averaged over 50 network realizations. In Figure 2, we illustrate the dynamical behavior of the information diffusion model which agrees with the theoretical prediction under normalized time span. A key feature of our model as compared to that of [16] is that known agents are allowed to be non-vanishing (c.f. Figure 2(b)) since known agents cannot become informed automatically, i.e., \( p_4 = 0 \) [16]. On the other hand, under the choice of parameters as
shown in Figure 2(a), we observed that all the agents have heard the information in the end, i.e. \( R(t) \) tends to \( N \) as \( t \) grows.

5 Concluding remarks

In this paper, analytical efforts are made to solve the affinity based information diffusion model introduced in [16]. Based on the time-inhomogeneous Markov chain characterization and the Wei-Norman method, exact solution for the model is obtained in terms of matrix exponentials. The related final-size equation is also recovered. While the application of Wei-Norman method in physical science has a long history [21, 25], its application in some simple biological population dynamics, such as SIR and SIS, appears only recently (see e.g. [24, 26–28]) due to lack of symmetry in such systems (Hence, it would be much more difficult to construct an appropriate Lie algebra with a low dimension). It is hoped that the approach offered in this study could shed some light on the analytical solution of more complicated (and realistic) social dynamics models.

As for possible extensions of the information diffusion model, we would like to mention two of them in addition to those discussed in [16]. Firstly, in the present framework, agents spread information but not affinity. In the real world, however, an agent might as well spread her affinity—she may be aware of her neighbors’ affinity and changes her own accordingly. Namely, both the message and the affinity with it may disseminate through the social network. Secondly, the only topological property of the substrate network considered here is the (average) degree. It would be appealing to incorporate some more informative quantities such as multi-rational structure [29] and social contagions with memory effect [30] in the information spreading mechanism.

Acknowledgement: This work is supported in part by the National Natural Science Foundation of China (11505127) and the Program for Young Excellent Talents in Tongji University (2014KJ036).

Conflict of Interests: The author declares that there is no conflict of interest regarding the publication of this paper.

References