SENSITIVITY TO NOISE VARIANCE IN A SOCIAL NETWORK DYNAMICS MODEL

BY

H. T. BANKS (Center for Research in Scientific Computation, Box 8205, North Carolina State University, Raleigh, North Carolina 27695–8205),

A. F. KARR (National Institute of Statistical Sciences, P.O. Box 14006, Research Triangle Park, North Carolina 27709–4006),

H. K. NGUYEN (Center for Research in Scientific Computation, Box 8205, North Carolina State University, Raleigh, North Carolina 27695–8205),

AND

J. R. SAMUELS, JR. (Center for Research in Scientific Computation, Box 8205, North Carolina State University, Raleigh, North Carolina 27695–8205)

Abstract. The dynamics of social networks are modeled with a system of continuous Stochastic Ordinary Differential Equations (SODE). With the proper amount of noise input, the SODE model captures dynamic features that are lacking in the corresponding deterministic ODE model. Therefore, sensitivity to noise levels is investigated by considering four different regimes: essentially deterministic, noise-enriched, noise-enlarged, and noise-dominated. Each regime is defined based on the behavior of solutions of the SODE, and geometry of the regimes are categorized with stochastic simulations.

1. Introduction. Social networks are a metaphor and mathematical construct for representing relationships among individuals.\footnote{Although the metaphor is the same, physical networks such as communication or transportation networks have quite different properties, and are modeled quite differently.} In a social network, nodes represent individuals, which we term agents, who may have both observable characteristics (for instance, employer or location) and unobservable characteristics (for instance, political or religious affiliation), while edges represent associations between pairs of agents. These associations may be uni-directional—leading to a directed graph, or bi-directional—leading to an undirected graph. Edges may have associated numerical values, representing for example the strength of the association.

Received November 18, 2005.
2000 Mathematics Subject Classification. Primary 91D30, 34F05, 60H10, 60H35, 90B15.
E-mail address: htbanks@ncsu.edu
E-mail address: karr@niss.org
E-mail address: hnguyen@ncsu.edu
E-mail address: jrsamue2@ncsu.edu

\(\text{c}2008\) Brown University
The underlying assumption in most social networks is that pairwise relationships are fundamental. Higher-order relationships are constructed from pairwise ones. Thus, for example, a \textit{triad} is a set of three agents in which each pair is joined by an edge (i.e., these agents define a complete subgraph). The usual social network framework does not seem to readily accommodate structures such as agents who are members of a committee but have no pairwise relationships.

There is a significant body of research on social networks, within which are threads addressing:

\textbf{Modeling:} often using probability models in which edges occur randomly as a function of the values of observable and unobservable characteristics, which are themselves often random.

\textbf{Visualization:} because some properties of nodes (for example, centrality—having an extremely large number of edges), even though defined mathematically, are strikingly apparent in good visualizations.\footnote{There is an extensive literature on graph visualization that is largely independent of the social network literature.}

\textbf{Inference:} for example, for parameters of network models.

\textbf{Prediction:} for example of characteristics of the entire network (e.g., average node connectivity) from a sample of the nodes and edges.

Quantitative studies of social networks have been recognized as important for many decades beginning with the work of Moreno \footnote{2} in 1934, Katz in 1947 and 1955 \footnote{15 \cite{16}}, and Festinger in 1949 and 1954 \footnote{7 \cite{8}}. Current research can be found in diverse areas including sexually transmitted diseases such as HIV, movie co-starring, needle sharing \footnote{12}, and terrorist networks \footnote{3 \cite{6}}, to name a few.

There are many types of models used to study social networks including the exponential family of social network models \footnote{9} or the $p^*$ class of models \footnote{26}, stochastic blockmodels \footnote{23}, intelligent agent models \footnote{2 \cite{5 \cite{14 \cite{22}}}}, and models with latent group structure \footnote{1 \cite{13 \cite{19}}}. For an overview of modern topics and approaches, we refer the reader to \footnote{4 \cite{24 \cite{25}}}. Most of the existing literature deals with static networks, in which neither the set of nodes nor the set of edges varies over time. By contrast, in this paper we introduce a (simple) evolution model of social networks that is motivated by the work of Hoff, et al., \footnote{13}. We represent the dynamics of relationships between individuals by a coupled system of nonlinear stochastic ordinary differential equations in which the pairwise relationships are driven by both observable and unobservable (latent) characteristics.

In our model we assume that time, the observable characteristics, and individual relationships are continuous. In the spirit of \footnote{13} and other models based on concepts of \textit{homophily}, we assume that an individual’s characteristics become more like those of other individuals to which it is linked. We also assume that links between nodes are more likely the higher the similarity of their characteristics. The dynamics (see \footnote{2}) are modeled using stochastic ordinary differential equations, in which the randomness represents unobservable factors not captured in the model. An alternative approach to modeling evolution of social network dynamics has been pursued by Snijders \footnote{21} using
Much of our attention focuses on qualitative behavior of the model as the noise parameters associated with stochastic disturbances vary. Heuristically, as the variance terms in the stochastic differential equations increase from 0, the model at first differs little from the deterministic case (in which they are 0). For somewhat larger values, the model differs little from the deterministic case (in which they are 0). For somewhat larger values, the model at first differs little from the deterministic case (in which they are 0). For still larger values of the variances, there may be noise-dominated behavior: a larger set of long-term behaviors is possible. Finally, for sufficiently large noise variances, the model is noise-enriched regime, in which its behavior is richer, but its ultimate fate remains that of the deterministic case. For still larger values of the variances, there may be noise-enlarged behavior: a larger set of long-term behaviors is possible. Finally, for sufficiently large noise variances, the model is noise-dominated, with no structure. Of course, the boundaries between these regimes are not sharp, and so we devote some attention simply to defining them. Nor is it clear that all exist in any specific situation.

2. Mathematical model. We denote by $C_i(t) \in K \subset \mathbb{R}^m$ the vector of characteristics of agent $i$, $i \in \mathcal{N} = \{1, \ldots, N\}$ at time $t$, where $m$ is the number of characteristics and $K$ is a (compact) constraint set for the values of the characteristics.

Let $e(i, i', t) \in \mathbb{R}$ be the degree of connectivity between agent $i$ and agent $i'$ at time $t$. Then a link or edge is present from agent $i$ to agent $i'$ at time $t$ if (and only if) $e(i, i', t) > 0$. By convention there are no "self links" so that $e(i, i, t) = 0$ for all $i$ and $t$. We do not require that connectivity be symmetric, so it is possible that $e(i, i', t) \neq e(i', i, t)$.

We let $A_i(t) = \{i' \in \mathcal{N} : e(i, i', t) > 0\}$ represent the set of all agents to which agent $i$ has links at time $t$ and let $|A_i(t)|$ be the number of elements in $A_i(t)$. Then the basic stochastic differential equations in our model are

$$dC_i(t) = \frac{\beta_i}{|A_i(t)|} \sum_{i' \in A_i(t)} [C_{i'}(t) - C_i(t)] \, dt + \sigma dW_{C_i}^e(t)$$

(2.1)

$$de(i, i', t) = f(\|C_i - C_{i'}\|) \, dt + \gamma dW_{e,i,i'}^e(t).$$

(2.2)

The coupling between (2.1) and (2.2) is via $A_i(t)$ since $A_i(t)$ depends on the vector $e_i(t) = (e(i, 1, t), \ldots, e(i, i', t), e(i, i' + 1, t), \ldots, e(i, N, t))$. In (2.1), $\beta_i > 0$ is an agent-specific factor that can be interpreted as the degree to which agent $i$ seeks to become more like other agents to which it is linked, and $\sigma W_i^C(\cdot)$ is a (vector-valued) Wiener process with variance $\sigma^2$. In (2.2), $f(\xi) = 2e^{-b\xi} - 1$, where $b > 0$, $\|\cdot\|$ denotes the norm in $\mathbb{R}^m$, and $\gamma W_{i, i'}^e(\cdot)$ is a Wiener process with variance $\gamma^2$.

Note that (2.1) actually consists of $Nm$ equations, while (2.2) consists of $N(N - 1)$ equations. These are to be solved subject to the state constraint $C_i(t) \in K$, with given initial conditions $C_i(0)$ and $e_i(0)$. All $Nm + N(N - 1)$ Wiener processes are assumed independent.

The interpretation of these equations is as described informally in §1. First, (2.1) stipulates that the characteristics of agent $i$ move at rate $\beta_i$ toward the centroid of the characteristics of the agents to which $i$ has links, subject to random perturbation from $\sigma W_i^C(\cdot)$. Then (2.1) states that the degree of connectivity between $i$ and $i'$ increases or decreases at rate $f(\|C_i - C_{i'}\|)$, again with a Wiener-driven perturbation.
The particular function $f$ in (2.1) is somewhat arbitrary. The important properties are that $f = 1$ when characteristics are identical, causing connectivity to increase at rate 1, and $f < 0$ when two agent’s characteristics are sufficiently unlike, causing their connectivity to decrease. The value for which $f = 0$, and hence characteristics have no effect on connectivity, depends on $b$. The impact of $b$ on formation of links is discussed in 5.1.

We note that (2.1) does not satisfy typical assumptions regarding smoothness of coefficients; an alternative, smoother model is described in 3.

3. Numerical method. In solving the above SODEs (2.1) and (2.2) subject to the constraints $C_i(t) \in K$, we use the stochastic analogue of the classical fourth-order Runge-Kutta method [10]. Although an implicit numerical scheme [17] would have more stability, we successfully used a familiar and simpler scheme to demonstrate model properties in our initial calculations. Before we write down the algorithms of the fourth-order Runge-Kutta method, we rewrite equation (2.1) as follows:

$$
dC_i(t) = g(t, C_i(t)) dt + \sigma dW^C(t),
$$

(3.1)

where $g(t, C_i(t)) = \frac{\mu_i}{|A_i(t)|} \sum_{i' \in A_i(t)} [C_{i'}(t) - C_i(t)]$. We partition the interval $[t_0, T]$ uniformly by the step size $h = \Delta t_n = t_n - t_{n-1}$. Then we can approximate $C_i(t_n)$ by the following:

$$
C_i(t_n) = C_i(t_{n-1}) + p_0 F_0 h + p_1 F_1 h + p_2 F_2 h + p_3 F_3 h + \sigma \Delta W^C_{in},
$$

(3.2)

where

$$
F_0 = g(t_{n-1}, C_i(t_{n-1}))
$$

$$
C_i^{(1)}(t_{n-1}) = C_i(t_{n-1}) + \frac{1}{2} F_0 h + \frac{1}{2} \Delta W^C_{in}
$$

$$
F_1 = g(t_{n-1} + \frac{1}{2} h, C_i^{(1)}(t_{n-1}))
$$

$$
C_i^{(2)}(t_{n-1}) = C_i(t_{n-1}) + \frac{1}{2} F_1 h + \frac{1}{2} \Delta W^C_{in}
$$

$$
F_2 = g(t_{n-1} + \frac{1}{2} h, C_i^{(2)}(t_{n-1}))
$$

$$
C_i^{(3)}(t_{n-1}) = C_i(t_{n-1}) + F_2 h + \Delta W^C_{in}
$$

$$
F_3 = g(t_{n-1} + h, C_i^{(3)}(t_{n-1}))
$$

$$
p_0 = p_3 = \frac{1}{6} \quad \text{and} \quad p_1 = p_2 = \frac{1}{3}.
$$

Here $\Delta W^C_{in} = W^C_i(t_n) - W^C_i(t_{n-1})$ are standard Wiener increments ([11], p. 489), which are obtained as sample values of normal random variables with the mean of zero and variance $\Delta t_n$. Equation (2.2) can be approximated in a similar manner, except it is simpler than equation (2.1) since the right-hand side of equation (2.2) is independent of the degree of connectivity $e(i, i', t)$.

The discretized system is solved subject to the constraint $C_i(t_n) \in K$. For any step resulting in a value violating this constraint, the value is replaced by the nearest value on the boundary of $K$. 

License or copyright restrictions may apply to redistribution; see http://www.ams.org/license/jour-dist-license.pdf
4. Numerical simulations. To illustrate the concepts introduced in §1, we employ an example with ten agents \((N = 10)\) and two characteristics \((m = 2)\) for each agent, each constrained to lie in \([-10, 10]\). Thus \(K = [-10, 10] \times [-10, 10]\). For concreteness, we term the first characteristic \(C_i^{sq}\) a “sociability quotient” (positive values represent higher sociability) and the second an “outlook on life” denoted by \(C_i^{out}\). For the latter, -10 is construed as a “negative outlook” and 10 a “positive outlook.”

For simplicity, we set \(\beta_i = 1\) for all \(i\), which in effect simply calibrates the time scale.

The initial values of the characteristics are set as \(C_2(0) = (-10, -10)\), \(C_7(0) = (-5, -5)\), and \(C_i(0) = (10, 10)\) for \(i \neq 2, 7\). We define the initial conditions for the connectivities \(e\) by \(e_i(0) = 0\).

4.1. Deterministic case. In the deterministic case that \(\gamma = \sigma = 0\), the behavior of the model depends (only) on the value of \(b \in (0, \infty)\). For \(b \in (0.105, \infty)\), neither agent \(A_2\), the “extreme outlier,” nor agent \(A_7\), the “moderate outlier,” forms edges with any other agents. The remaining agents, of course, form edges and their connectivity continues to increase. This behavior can be seen by examining the degrees of connectivity \(e(2, i', T), e(7, i', T),\) and \(e(10, i', T)\) (throughout, we use \(A_{10}\) as an exemplar of all agents other than \(A_2\) and \(A_7\)) at \(T = 10\). We illustrate in Figures 1, 2, and 3 this three-cluster scenario. The plots satisfy \(e(2, i', T) < 0\), \(e(7, i', T) < 0\) for all \(i'\), and \(e(10, i', T) < 0\) for \(i' = 2, 7\), while \(e(10, i', T) > 0\) for \(i' \neq 2, 7\).

Specifically, in Figure 1 the connectivity between \(A_2\) and the other nine agents is negative at \(T = 10\). In other words, \(A_2\) does not have links to any other agent at the final time. We see the same result for \(A_7\) in Figure 2. Finally, Figure 3 shows that \(A_{10}\) has no links (negative connectivity) with \(A_2\) and \(A_7\), but links (positive connectivity) to all other agents.

A two-cluster regime arises for \(b \in [0.033, 0.105]\). One cluster contains \(A_2\) and \(A_7\), and the other cluster consists of the remaining agents. As we illustrate in Figure 4, \(A_2\) becomes linked to \(A_7\) by \(T = 10\), but not to any other agents. Figure 5 shows that \(A_7\) is linked to \(A_2\) but not to any of the remaining agents. Figure 6 shows that \(A_{10}\) has links to all agents except \(A_2\) and \(A_7\).

Finally, for \(b \in (0, 0.033)\), we have a one-cluster case: all ten agents become linked to one another. By observing the connectivity of \(A_2\) in Figure 7, \(A_7\) in Figure 8, and \(A_{10}\) in Figure 9, we note that \(A_7\) is more strongly connected to the group than \(A_2\) (as expected since \(A_7\) is the “moderate outlier”). Moreover, \(A_7\) and \(A_2\) have higher connectivity to one another than with the remaining agents.
FIG. 1. Degree of connectivity $e(2, i', t)$ of agent 2 (the “extreme outlier”) with other agents for $b = 0.15$. Note that the curves for $i' = 1, 3, 6, 8, 10$ all lie on the same line in this 3-cluster scenario.

FIG. 2. Degree of connectivity $e(7, i', t)$ of agent 7 (the “moderate outlier”) with other agents for $b = 0.15$ (again the 3-cluster scenario).
Fig. 3. Degree of connectivity $e(10, i', t)$ of agent 10 with other agents (except for 2 and 7) for $b = 0.15$ (again a 3-cluster scenario).

Fig. 4. Degree of connectivity $e(2, i', t)$ of agent 2 with other agents for $b = 0.06$ illustrating the two-cluster scenario.
Fig. 5. Degree of connectivity $e(7,i',t)$ of agent 7 with other agents for $b = 0.06$ (again the 2-cluster scenario).

Fig. 6. Degree of connectivity $e(10,i',t)$ of agent 10 with other agents for $b = 0.06$ (2-cluster scenario).
Fig. 7. Degree of connectivity $e(2,i',t)$ of agent 2 with other agents for $b = 0.0165$ illustrating the single-cluster scenario.

Fig. 8. Degree of connectivity $e(7,i',t)$ of agent 7 with other agents for $b = 0.0165$ (single-cluster scenario).
4.2. Stochastic case: Regime definition. We categorize pairs $(\sigma, \gamma)$ based on the behavior of solutions of the SODEs (2.1) and (2.2) with fixed $b = 0.0165$. Recall from §4.1 that in this case the deterministic system leads to a single cluster, where A7 has a higher connectivity with other agents than A2, and A7 and A2 are more connected to one another than to other agents for all time $t$.

With the stochastic terms included (that is, $\gamma > 0$ and $\sigma > 0$), we expect a broader range of behavior for the solutions to (2.1) and (2.2). In §1 we defined four regimes heuristically. For the example, we formalize those definitions as follows. The final time is $T = 30$.

**Essentially deterministic:** consists of $(\sigma, \gamma)$ pairs for which all of the following are satisfied:

- **ED1:** For all $t \leq T$, all connectivities between all agents are nonnegative: $e(i, i', t) \geq 0$, $0 \leq t \leq T$, $i \neq i'$.
- **ED2:** For all $t \leq T$, A7 and A2 have higher connectivity with each other than the rest of the group: $e(2, 7, t) > e(2, i', t)$ for $i' \neq 7$ and $e(7, 2, t) > e(7, i', t)$ for $i' \neq 2$.
- **ED3:** For all $t \leq T$, A7 has higher connectivity with others than does A2: $e(i, 7, t) > e(i, 2, t)$ for $i \neq 2, 7$.
- **ED4:** At $t = T$, all agents have links to one another: $e(i, i', T) > 0$ for $i \neq i'$. That is, in the essentially deterministic regime, all qualitative aspects of the deterministic case are preserved.
Noise-enriched: consists of $(\sigma, \gamma)$ pairs for which any of ED1–ED3 fails, but ED4 holds. That is, the agents take different paths to reach the final destination of being linked to one another at $T$, so that the social network evolves to a single cluster.

Noise-enlarged: consists of $(\sigma, \gamma)$ pairs for which ED4 also fails, but either the three-cluster or two-cluster “equilibrium” in §4.1 arises. Put differently, $(\sigma, \gamma)$ is in the noise-enlarged regime if either of the following holds:

- NE1: At time $T$, $A_7$ and $A_2$ are linked in one group and the remaining agents are linked in another group (two clusters).
- NE2: At time $T$, $A_7$ and $A_2$ are not linked to any other agents, but all agents other than $A_2$ and $A_7$ are linked to one another (three clusters).

Noise dominated: consists of $(\sigma, \gamma)$ pairs for which none of the three structured outcomes described above obtains.

To investigate whether these regimes exist and, if so, to characterize their properties, we ran simulations, with time step $h = \Delta t_n = 0.05$, for selected values of $\sigma \in [0, 2.7]$ and $\gamma \in [0, .8]$. We carried out 100 simulations for each pair $(\sigma_k, \gamma_l)$ in a partition with values $\sigma_k = .1k, k = 1, \ldots, 27$, $\gamma_l = .01l, l = 1, \ldots, 80$.

We depict the surfaces representing the noise-enriched and noise-dominated regimes in Figures 10 and 11, respectively. The noise-dominated surface is almost the complement of the noise-enriched surface. A few cases of essentially deterministic behavior are seen at small values of $\sigma$ and $\gamma$.

Therefore, we used the refined partition $\sigma_k = .01k, k = 1, \ldots, 70, \gamma_l = .001l, l = 1, \ldots, 50$ to attempt to delineate the essentially deterministic regime. The corresponding essentially deterministic and noise-enriched surfaces are plotted in Figures 12 and 13, respectively. These two plots are almost complements of one another because there are few cases of noise domination.

5. Concluding remarks. We observe from the simulations reported in §4.2 that regimes appear to be more sensitive to $\gamma$ than to $\sigma$. This may be because $\gamma$ directly affects edge formation, which is critical in regime definition, whereas $\sigma$ does not. We suspect that if the regime definitions were based upon the observed characteristics, which are directly influenced by $\sigma$, then the sensitivities observed might be reversed.

There is no evidence in our simulations for existence of a noise-enlarged regime. We believe this is because the criteria that define the noise-enlarged state depend on a nonlinear term involving $b$, whereas the randomness $\sigma$ and $\gamma$ enter the dynamics linearly. In order to verify this claim, we would need to pursue a rigorous sensitivity analysis for the SODEs (2.1) and (2.2).

Although we have a simple model, the richness and flexibility of the model are substantial due to the model parameters. For example, we can significantly enrich the model by allowing the parameter $b$ in the connectivity formation function $f$ to be agent-dependent. Moreover, we might have $\beta_i \in \mathbb{R}$ instead of $\beta_i > 0$ to describe situations where agent $i$ is either repelled by or attracted to the characteristics of his acquaintances. We would then expect the model to exhibit more than just the three regime cases described above. Since exhaustive simulation studies are not particularly desirable (nor feasible as the
Fig. 10. The \((\sigma, \gamma)\) noise-enriched surface plotted as a percentage or number out of 100 simulations.

Fig. 11. The \((\sigma, \gamma)\) noise-dominated surface.
Fig. 12. The \((\sigma, \gamma)\) essentially deterministic surface.

Fig. 13. The \((\sigma, \gamma)\) noise-enriched surface.
model complexity increases), development of a theoretical and computational framework for sensitivity analysis of such models should prove most useful. Our preliminary efforts and results reported on here provide a sound foundation for further investigations in this direction.

As noted in §2, (2.1) fails to have smooth coefficients because \( |A_i(\cdot)| \) changes discontinuously. Therefore, the solution is not a diffusion process in the usual sense. A smooth alternative to (2.1) is

\[
dC_i(t) = \frac{\beta_i}{\sum_{i' \neq i} e(i, i', t)} \sum_{i' \neq i} e(i, i', t) \left[ C_i(t) - C_{i'}(t) \right] dt + \sigma dW_i(t). \tag{5.1}
\]

In (5.1) the connectivities serve as weights. When combined with (2.2), (5.1) produces a smooth system in which—interestingly—edges or links have been replaced by smooth connectivities. More complete investigation of this system is a topic for future research.

Acknowledgements. This research was supported in part by the National Science Foundation under grant DMS–0437183 to the National Institute of Statistical Sciences (NISS), the U.S. Air Force Office of Scientific Research under grant AFOSR-FA9550-04-1-0220, and the National Science Foundation under grant DMS–0112069 to the Statistical and Applied Mathematical Sciences Institute (SAMSI). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. We wish to express our gratitude to Dr. Nathan Gibson for his assistance in parallel computing and to Dr. Michael Last for suggestions related to the regime definitions.

References


