1. Preliminaries

We consider a network with two types of agents, *people* and *influencers* (Figure 1.1). People change their opinions based on the opinions of other people and influencers. Influencers are insensitive to the opinions of others.

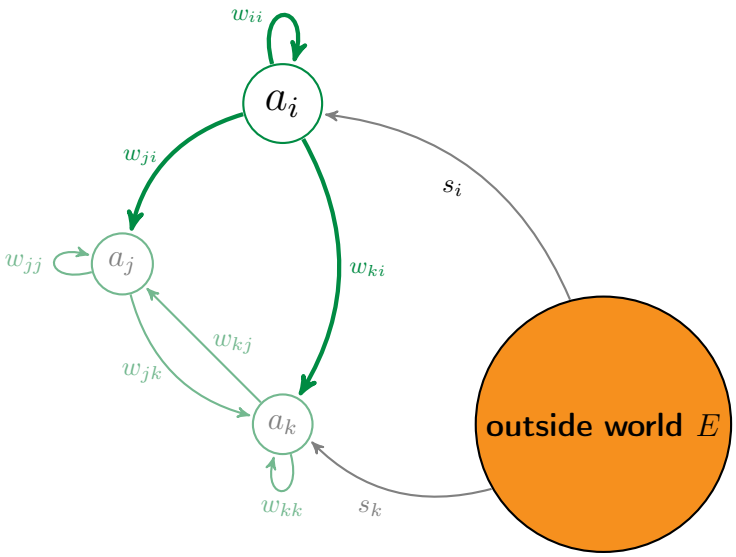

FIGURE 1.1. **Schematic of agent-based model of a social network**. The large burnt orange circle denotes input from the outside world to agents in the social network. Smaller green circles denote agents. Thicker and darker edges denote connections from an influencer. Thinner and lighter edges denotes connections from the outside world or people.

Equation 1.1 describes the activity of the ith agent at some time t, $a_i(t)$ in terms of that agent's prior activity, the activity of others in the network, and external influences. Table 1 describes the other variables in Equation 1.1.

$$\tau_i \frac{\mathrm{d}}{\mathrm{dt}} a_i(t) = -a_i(t) + \sum_{j \neq i} w_{ij} a_j(t) + s_i E(t) \tag{1.1}$$

Variable	Meaning
τ_i	Stubbornness, half-life at which $a_i(t) \to 0$ in the absence of other input
w_{ij}	Importance the ith user gives to the jth user's activity
s_i	Sensitivity of ith user to external input
$E(t)$	Strength of external input at time t

TABLE 1. Description of variables in Equation 1.1.

Table 2 describes the ranges of terms in Equation 1.1 that correspond to people and influencers.

Variable	Influencer	Person
τ_i	High	Low
w_{ij}	Zero	Nonzero
s_i	Zero	Nonzero

TABLE 2. Description of variables in Equation 1.1.

The last two terms on the right hand side of Equation 1.1 may be regarded as the first terms in Taylor expansions of two functions that describe the interactions between the ith user and the social network, f, and the external world, g (Equation 1.2).

$$\begin{aligned} f(a_i, a_j) &= \sum_j w_{ij} a_j + \sum_{j,k} w_{ijk} a_j a_k + \sum_{j,k,l} w_{ijkl} a_j a_k a_l \\ g(a_i) &= s_i E(t) + s_{i,2} E^2(t) + \ldots = \sum_n s_{i,n} E^n(t) \end{aligned} \tag{1.2}$$

Recognizing this interpretation of the last two terms underscores that (i) Equation 1.1 accounts only for *pairwise linear* interactions, and (ii) the model is readily extensible. We do not consider the higher-order terms in Equation 1.2 further in this paper.

2. Computational Implementation

In this section we discuss a computational implementation of the network as one of stochastic binary units (Boltzmann machine). In the Boltzmann machine, the activity of each user takes on one of two values (Equation 2.1). In the context of a social network, we take *active* to mean posting a message on that topic. Each agent would have a different activity function for each topic. Here, we only consider the activity of agents on one topic.

$$a_i(t) = \begin{cases} 1 & \text{active} \\ -1 & \text{inactive} \end{cases} \tag{2.1}$$

Equation 2.2 defines the state of the ith unit at some time t, $\sigma_i(t)$ in analogy with Equation 1.1.

$$\sigma_i(t) = \sum_j w_{ij} a_j + s_i E(t) \tag{2.2}$$

To simulate the dynamics of the network we select a single unit at random at each time step and update that unit according to Equation 2.3.

$$\mathbb{P}[a_i(t+1) = 1] = \frac{1}{1 + e^{-\sigma_i(t)}} \tag{2.3}$$

Because our network follows Equation 2.3, we may calculate how stable a pattern of activity, $\mathbf{a}$, is by defining the energy function ψ (Equation 2.4).

$$\psi(\mathbf{a}) = -\left(E(t) \sum_i s_i + \frac{1}{2} \mathbf{a} \cdot \mathbf{W} \cdot \mathbf{a} \right) \tag{2.4}$$

In Equation 2.4, the ijth entry of the matrix $\mathbf{W}$ is w_{ij} as introduced in Equation 1.1 and described in Table 1. Similarly, the ith member of the vector $\mathbf{a}$ is a_i, as introduced in Equation 1.1. From Equation 2.4 we can calculate how likely a pattern of activity, $\mathbf{a}$, is to occur in a social network where agents weigh each others activities according to $\mathbf{W}$ (Equation 2.5).

$$\mathbb{P}[\mathbf{a}] = \frac{e^{-\psi(\mathbf{a})}}{\sum_{\mathbf{a}} e^{-\psi(\mathbf{a})}} \tag{2.5}$$

Equations 2.4 and 2.5 are important because they allow us to calcualte how changes in the strength of interactions between agents, $\mathbf{W}$, change which patterns of activity dominate the social network. They allow us to rank networks as to their likelihood of exhibiting concerning patterns of activity. Finally, they allow us to identify which interactions among users contribute most to a concerning pattern of activity, which allows us to target specific nodes in the network.

Estimation of parameters. From Twitter we can reconstruct two types of social networks, depending on what the connections between users (edges between nodes) denote. In one schema, the *follower graph, a* connection runs from the ith agent to the jth agent if the ith agent follows the jth agent. In the other schema, a *retweet graph*, a connection runs from the ith agent to the jth agent if the ith agent retweets a tweet of the jth agent. The follower graph is directed but unweighted because no agent can follow another agent multiple times. The retweet graph, in contrast, is directed and weighted because an agent can retweet another agent's tweets many times.

We consider only follower graphs. Retweet graphs seem to be more stable (VERIFY THIS.) and may reflect more substantive connections.

3. Mathematical Analysis

Different agents will populate social networks on different topics. It is, accordingly, reasonable to focus on the patterns of interactions among users, because different groups of users may show similar patterns of interactions. Instead of directly analyzing the connection matrix, $\mathbf{W}$, we analyze the Laplacian of the connection matrix, $\mathcal{L}$ (Equation 3.1).

$$\mathcal{L}_{ij} = \begin{cases} 1 - \frac{w_{ij}}{dj} & i = j \\ -\frac{w_{ij}}{\sqrt{d_i d_j}} & i, j \text{ adjacent} \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

In Equation 3.1, the symbol d_j denotes the degree of the jth node, the number of agents that follow agent j or are followed by agent j. Laplacian matrices are symmetric and positive semi-definite. Equation 3.2 expresses the activity of the network in terms of the eigenvalues of the Laplacian of the connection matrix.

Equation 3.2 re-expresses Equation 1.1 in terms of the μ eigenvectors of the connection matrix, $\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_\mu$.

$$\mathbf{a}(t) = \sum_{\mu=1} c_\mu(t)\,\mathbf{e}_\mu \tag{3.2}$$

Equation 3.3 uses the property that all eigenvectors are mutually orthogonal, $\mathbf{e}_\mu \cdot \mathbf{e}_\nu = \delta_{\mu\nu}$, to describe the evolution of the coefficients in Equation 3.2.

$$\frac{\mathrm{d}}{d\mathrm{t}} c_\nu(t) = -(1-\lambda_\nu)\,c_\nu(t) + E(t)\cdot\mathbf{e}_\nu \tag{3.3}$$

The spectrum (distribution of eigenvalues) uniquely identifies a pattern of connections. If the Laplacian of two networks share the same spectrum, then the patterns of connection within them are functionally indistinguishable and information flows through those networks in the same way.

4. Validation