

# Latent Space Diffusion

by

Jacob C. Fisher

Department of Statistical Science  
Duke University

Date: \_\_\_\_\_

Approved:

---

David Banks, Supervisor

---

James W. Moody

---

Katherine A. Heller

Thesis submitted in partial fulfillment of the requirements for the degree of  
Master of Science in the Department of Statistical Science  
in the Graduate School of Duke University  
2015

ABSTRACT

Latent Space Diffusion

by

Jacob C. Fisher

Department of Statistical Science  
Duke University

Date: \_\_\_\_\_

Approved:

---

David Banks, Supervisor

---

James W. Moody

---

Katherine A. Heller

An abstract of a thesis submitted in partial fulfillment of the requirements for  
the degree of Master of Science in the Department of Statistical Science  
in the Graduate School of Duke University  
2015

Copyright © 2015 by Jacob C. Fisher  
All rights reserved except the rights granted by the  
Creative Commons Attribution-Noncommercial Licence

# Abstract

Social networks represent two different facets of social life: (1) stable paths for diffusion, or the spread of something through a connected population, and (2) random draws from an underlying social space, which indicate the relative positions of the people in the network to one another. The dual nature of networks creates a challenge - if the observed network ties are a single random draw, is it realistic to expect that diffusion only follows the observed network ties? This study takes a first step towards integrating these two perspectives by introducing a social space diffusion model. In the model, network ties indicate positions in social space, and diffusion occurs proportionally to distance in social space. Practically, the simulation occurs in two parts: positions are estimated using a latent space model, and then the predicted probabilities of a tie from that model - representing the distances in social space - or a series of networks drawn from those probabilities - representing routine churn in the network - are used as weights in a weighted averaging framework. Using a school friendship network, I show that the model is more consistent and, when probabilities are used, the model converges faster than diffusion following only the observed network ties.

To Mom, Dad, and Jane. I couldn't have done any of this without your love and support.

# Contents

<b>Abstract</b>	<b>iv</b>
<b>List of Figures</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background</b>	<b>3</b>
2.1 Diffusion models on changing networks . . . . .	3
2.2 Latent space models for networks . . . . .	6
2.3 The weighted averaging diffusion model . . . . .	8
2.4 Incorporating the latent space model into the weighted averaging process	10
<b>3 Data</b>	<b>14</b>
3.1 Sample . . . . .	14
3.2 Latent space model specification . . . . .	15
<b>4 Results</b>	<b>16</b>
<b>5 Discussion</b>	<b>25</b>
<b>Bibliography</b>	<b>28</b>

# List of Figures

4.1	Visualization of weighted averaging diffusion processes . . . . .	18
4.2	Change in simulated attitudes after each iteration . . . . .	20
4.3	Distribution of attitude values by simulated iteration . . . . .	21
4.4	Mean squared error of predictions . . . . .	23

# Acknowledgements

The author thanks Jim Moody, David Banks, Katherine Heller, Esther Sackett, Jeff Smith, Robin Gauthier, and members of working groups at Duke University and Pennsylvania State University for their helpful comments. Grants from the W.T. Grant Foundation (8316), the National Institute on Drug Abuse (R01-DA08225), and an NIA training grant to the Center for Population Health and Aging at Duke University (T32 AG000139) supported this research. The analyses used data from PROSPER, a project directed by R. L. Spoth, funded by grant R01-DA013709 from the National Institute on Drug Abuse and co-funded by the National Institute on Alcohol Abuse and Alcoholism (grant AA14702).

# 1

## Introduction

Social networks have historically been used for two different purposes. First, social networks represent the paths that contagions - like innovations such as the use of a new medicine (Coleman et al., 1966), information such as job opportunities (Granovetter, 1973, 1995), diseases such as HIV (Morris et al., 2009), or even a medical condition such as obesity (Christakis and Fowler, 2007) - take to spread through a population. Second, social networks represent the relative positions of people - for example, status (Rossman et al., 2010), hierarchy (Martin, 2002), popularity (Moody et al., 2011), informal peer group membership (Newman, 2006), role structures (White et al., 1976), or levels of intergroup contact (Smith et al., 2014) - to one another. Under the former framework, network ties indicate stable conduits for social influence or information; under the latter, network ties are simply a random draw from an underlying social space.

The dual nature of networks as stable connections and as random draws from social space creates a challenge - if the observed network ties are a single random draw, is it realistic to expect that diffusion, or the spread of something through a connected group of people, only follows the observed network ties? Under experimental

conditions, where online network ties are created by the researcher (e.g., Centola, 2011), it is reasonable to assume that diffusion can only follow the given network ties. A growing literature on measurement error in networks (Eagle and Proeschold-Bell, 2015; Paik and Sanchagrin, 2013), however, suggests that in observed networks, the specific ties that are observed depends heavily on how the data were collected. As such, in real networks, the absence of a tie may not be meaningful, and it may not be reasonable to assume that there is no direct information transmission when no tie is observed.

To address this challenge, I introduce a new class of network diffusion models, social space diffusion models. In a social space diffusion model, ties represent a random draw from an underlying social space, and the spread of a contagion between two people occurs proportionally to distance in social space, rather than strictly following the presence or absence of a tie. To estimate distances in social space, I use a latent space model (Hoff et al., 2002; Handcock et al., 2007), and then I simulate diffusion over those distances using the weighted averaging model elaborated by Friedkin and colleagues (Friedkin, 1998; Friedkin and Johnsen, 2011). I illustrate the method using data a friendship network in an American middle school, to show how the results from a social space diffusion model differs from the weighted averaging model using the observed network. Although I focus on the weighted averaging model for analytical tractability, this method, with appropriate modifications, could be extended to other diffusion models, such as threshold models (Centola and Macy, 2007; Granovetter, 1978; Watts, 2002) or voter models (Durrett et al., 2012; Holme and Newman, 2006).

# 2

## Background

### 2.1 Diffusion models on changing networks

The two approaches to networks, networks as stable connections that carry information and influence, and networks as fluid indicators of relative positions, can be considered in terms of the simplifying assumptions made by researchers modeling the changes in the network over time. Most important of these is the choice to hold either the network or the diffusing bit constant, based on which one is changing faster.

When the network changes relatively slowly compared to the thing that is spreading through the network, holding the network constant and studying the movement of a bit through the network can be a reasonable approximation. Modeling bits spreading through a static network is sometimes called a social influence process or dynamics on the network. For example, in studies of attitude diffusion, Friedkin and colleagues Friedkin and Johnsen (2011) generally assume that the network is unchanging while people influence each other. The network may change, but it is changing slowly enough that it can be approximated by a static network. Holding

the network constant treats the network ties as stable conduits for a diffusing piece of information to travel across.

By contrast, when the network changes relatively quickly compared to the thing that is spreading, studying changes in the network, instead of changes in the thing spreading through the network, becomes a better approximation. For example, in the Schelling model of racial segregation in housing Schelling (1971, 1978), race does not change over each person's life, but people can change their neighbors. The Schelling model focuses on the change in network ties - meaning neighbors in this case - and assumes that the thing spreading through the network, race, is not moving between people at all. Models that examine the change in network and hold spreading bits constant are referred to as homophily processes or as models of dynamics of networks. Holding the diffusing bit stable treats network ties as fluid indicators of how much people who share a particular attribute interact, or as the relative social position of the people holding one attribute to people holding the other.

A set of hybrid models exist, for when the network changes at roughly the same time scale as the thing spreading through it and neither the network nor the thing spreading through it can be held stable as an approximation. Voter models (Durrett et al., 2012; Holme and Newman, 2006), for example, consider changes in the network and changes of the attitudes of people in the network simultaneously; at every time step, a person has the option of changing his or her mind, or changing who he or she is connected to. Studies where the network changes and the diffusing bit coevolve model selection and influence processes and may be referred to as dynamics of and on the network.

Each of these approximations, however, treats changes in the network as meaningful variation, rather than stochastic. Recent research on network measurement, however, has shown that measuring networks introduces a considerable amount of stochastic variation. Paik and Sanchagrin (2013), for example, showed that signif-

icant differences in core discussion networks could be traced back to differences in which interviewer contacted the respondent. Eagle and Proeschold-Bell (2015) similarly showed that surveys administered by web and by phone produced differently sized core discussion networks. Additionally, Eagle and Proeschold-Bell found evidence for interviewer effects, as Paik and Sanchagrin found, and panel conditioning, meaning that if a person had taken the survey before, he or she would report a smaller network on the next round of the survey, presumably to avoid answering as many follow-up questions.

In addition to systematic variation caused by how the network ties were elicited, network ties also show random, routine churn. Considering two surveys of adolescent friendships separated by three weeks, for example, Cairns and Cairns Cairns and Cairns (1994) found a 0.74 probability that a friendship would remain. Although friendship churn among adolescents appears to be associated with some changes in the underlying social structure of the school (e.g., Moody et al., 2011), much of the change is idiosyncratic, simply representing measurement error in the friendship ties.

Stochastic variation in the network ties introduces another possibility for diffusion models. Changes in network ties can be decomposed into two parts, systematic variation and random variation. The two types of variation may occur at different rates. The random variation, meaning the routine churn in network ties, may occur at roughly the same rate as the movement of the thing diffusing through the network, while the systematic variation may occur much more slowly. That is, the actual friendships that a researcher would observe may change on a day to day basis, but the underlying social structure, such as the peer groups, the students' academic tracks, or the rates of interaction between sociodemographic groups may be much more stable. Under those circumstances, although the network's stochastic changes happen at roughly the same rate as the spread of the thing diffusing through the network, the systematic structure of the network changes much more slowly than the

spread of the diffusing thing.

In that case, an appropriate network diffusion model could hold the underlying social structure stable and model diffusion of something over that space, while accounting for the stochastic changes in a network. Thus a model would proceed by using the observed network to estimate the positions of the people in an underlying social space. The model could then estimate diffusion over various realizations of the social space, or could model diffusion over the social space directly. To develop such a model, I use the latent space model framework (Handcock et al., 2007; Hoff et al., 2002)<sup>1</sup>, described in the following section.

## 2.2 Latent space models for networks

Latent space models were introduced by Hoff et al. (2002), and have since been expanded to include model-based clustering (Handcock et al., 2007) and dynamic networks (Sewell and Chen, 2015). Latent space models are similar to a logistic regression predicting whether or not a tie will occur between each pair of people in the network. The models include a random effect - a position in the latent space - for every person. The latent positions are usually constrained to lie in a low-dimensional, Euclidean space to make the model easier to fit and to interpret. A tie is more likely between 2 people who are closer in the latent space.

More formally, for a network with  $n$  people, or vertices, let  $Y$  denote the  $n \times n$  adjacency matrix, whose elements,  $y_{i,j}$ , are 1 if  $i$  is connected to  $j$ , and are 0 otherwise. The latent space model treats the connections, which I will also refer to as ties or edges, as conditionally independent given the vertices positions in an underlying

<sup>1</sup> The exponential-family random graph model (ERGM) is a common, and perhaps more popular, alternative to the latent space model framework. Both approaches attempt to control for common network effects like transitivity and reciprocity; ERGMs explicitly parameterize those effects, while latent space models control for those effects implicitly by embedding vertices in a latent space. Since I am only interested in obtaining estimates for the probability of two people being connected, and not in modeling the processes that caused them to be connected, I use the latent space model framework.

latent space. That is,

$$Pr(Y | X, B, Z) = \prod_{i \neq j} P(y_{i,j} | x_{i,j}, z_i, z_j, \theta) \quad (2.1)$$

Where  $X$  is the matrix of covariates on each pair of vertices with elements  $x_{i,j}$ ,  $Z$  is the matrix of latent positions and  $\theta$  are a set of parameters. Both  $Z$  and  $\theta$  are to be estimated from the data. The most common parameterization is a logistic regression framework where the latent positions are constrained to lie in a low-dimensional, Euclidean space. The model is then written as

$$\eta_{i,j} = \text{log-odds}(y_{i,j} = 1 | x_{i,j}, z_i, z_j, \beta) \quad (2.2a)$$

$$= \beta^T x_{i,j} - |z_i - z_j|. \quad (2.2b)$$

That is, the log-odds of a tie is proportional to the distance of two vertices in the latent space. Since the positions themselves are not identified, the  $|z_i - z_j|$  term is replaced with a single distance term,  $\delta_{i,j}$ . Although any distance that satisfies the triangle inequality can be used (e.g., Hoff, 2005, 2009), usually distances are modeled in a low-dimensional Euclidean space for parsimony and ease of interpretation.

People who are isolated, meaning they do not have any ties to others in the observed data, present a problem for latent space models, since the maximum likelihood estimate of the distance from an isolate to any other person in the network is infinite. To address this problem, most authors fit latent space models in a Bayesian framework (Handcock et al., 2007; Krivitsky and Handcock, 2008; Sewell and Chen, 2015). Giving the latent positions a minimally informative prior, centered at 0, shrinks the position of isolates towards the center of the latent space, which gives isolates a positive probability of being connected to others in the network. I also take a Bayesian approach to estimation, using the minimally informative priors and the MCMC algorithm outlined by Krivitsky and Handcock (2008). I denote draws from the posterior distribution of the parameters as  $\theta^{(1)}, \dots, \theta^{(K)}$ , and draws from

the posterior predictive distribution as  $Y^{(1)}, \dots, Y^{(K)}$ .

Under the latent space framework, ties are modeled as random realizations of an underlying latent space. Thus changes in observed ties represent stochastic, rather than systematic, variation in the distances between the vertices. Viewing the distances as a type of social distance follows the positionist approach to social networks (Burt, 1987), where stable social distances govern the process of tie formation and dissolution (McPherson, 1983; McPherson et al., 2001; Smith et al., 2014). While typical diffusion approaches, called connectionist approaches, treat the ties as stable conduits for information, a positionist approach treats ties as indications of the social position of each person relative to the other people in the network.

The connectionist and positionist approaches generally have been viewed as mutually incompatible. If a tie is a random draw from an underlying social space, then it cannot also be a stable conduit for information. Viewing ties as random realizations of an underlying social space provides several benefits for modeling diffusion, however. First, once fit, the model incorporates missing data gracefully. Since each of the predicted probabilities of a tie is conditionally independent given the distances, if one vertex is missing, it will not affect the predicted probability of a tie between any other pair of vertices. Second, the model predictions incorporate stochastic variation in the ties. The stochastic variation can be incorporated in one of two ways: either by using the predicted probabilities of a tie or by using networks drawn from those probabilities (that is, the predictive distribution), to model diffusion. The following section introduces a simple diffusion model and describes these two approaches in the context of that model in greater detail.

### 2.3 The weighted averaging diffusion model

For this study, I use a weighted averaging diffusion model to illustrate how a latent space model could be used in conjunction with a diffusion model. With appropriate

modifications, however, a latent space approach could be used with other diffusion models, such as the Watts threshold model (Watts, 2002) or a complex contagion model (Centola and Macy, 2007). The weighted averaging model developed independently in sociology (French, 1956; Harary, 1959; Friedkin, 1998; Friedkin and Johnsen, 2011) as a model for the diffusion of attitudes, and in statistics (DeGroot, 1974; Chatterjee and Seneta, 1977; Berger, 1981) as a model for the formation of consensus. At its core, the model supposes that each person in a group starts with a particular attitude about a given topic, usually meaning that each person falls somewhere along the continuum from favorable to unfavorable for the topic. Then, under the model, each person updates his or her attitude by taking the average of his or her friends' attitudes. If this updating is done repeatedly, the group will eventually reach a consensus on the topic, under certain conditions.

Formally, the model is written as  $A^{(t+1)} = WA^{(t)}$ , where  $A^{(t)}$  is an  $n \times 1$  vector of people's opinions at time  $t$ , and  $W$  is an  $n \times n$  matrix whose elements,  $w_{i,j}$ , indicate the weight that person  $i$  gives to the opinion of person  $j$ . The weights are normalized such that the rows sum to 1, i.e.,  $\sum_{j=1}^n w_{i,j} = 1$  for each  $i = 1, \dots, n$ . Typically, the weights are given by the row-normalized adjacency matrix. Denoting the adjacency matrix as  $Y$ , typically  $w_{i,j} = y_{i,j} / \sum_{j=1}^n y_{i,j}$ . The diagonal of the matrix,  $w_{i,i}$ , represents the weight that a person places on his or her own attitude. Friedkin and Johnsen (2011) suggest a framework for choosing values for  $w_{i,i}$ ;<sup>2</sup> for simplicity, in this application I set all the diagonal elements equal to 1 before row normalizing the weights. Setting the diagonal elements to 1 means that each person gives his or her own opinion the same weight as 1 of his or her network neighbors.

The row normalized weight matrix can be thought of a transition matrix for a Markov chain. If the weight matrix does not change between iterations, the diffusion

---

<sup>2</sup> Specifically, Friedkin and Johnsen (2011) suggest the model  $A^{(t)} = BWA^{(t-1)} + (I - B)A^{(0)}$ , where  $B$  is a diagonal matrix of self-weights, and  $I$  is the identity matrix. The elements of  $B$ ,  $b_{i,i}$ , are set such that  $b_{i,i} = 1 - w_{i,i} = \sum_{j \neq i} w_{i,j}$ .

process at any time step can be written in terms of change from the initial position as  $A^{(t+1)} = W^t A^{(1)}$ , where  $A^{(1)}$  denotes the initial attitudes for each person. After many iterations, meaning as  $t \rightarrow \infty$ , the chain will stabilize into a stationary distribution, denoted  $W^\infty$ . This final distribution, which I denote  $A^{(\infty)} = W^\infty A^{(1)}$ , represents the value of the consensus that the group reaches. If the transition matrix for the chain representing the entire network is irreducible and aperiodic,<sup>3</sup> then everyone in the group will reach a single consensus value, where each person's contribution to the final consensus is given by their weight in the stationary distribution. If the transition matrix is not irreducible, as is the case for networks that have more than one strongly connected component, such as in the example to follow, there will be multiple consensus values, one for each strongly connected component in the network.

## 2.4 Incorporating the latent space model into the weighted averaging process

Using results from the latent space model, stochastic changes in the ties can be incorporated by changing the weight matrix  $W$ .  $W$  can either be changed to the predicted probability of a tie between each person  $i$  and  $j$ , or it can be changed to a simulated network drawn from the predictive distribution of the latent space model. Setting the elements of  $W$  as the predicted probability of a tie between each pair of people suggests that, at each iteration, a person takes the opinion of every other person in the network into account. People who are very distant still influence each other, but, because they are very distant, the probability of a tie between them is very low and they therefore give each other's attitudes very little weight. Formally, to set the elements of  $W$  to predicted probabilities, I transform the draws from

---

<sup>3</sup> Irreducible means that each person can get to each other person in the network; aperiodic means that the network does not have regular cycles that would cause the chain to oscillate back and forth indefinitely. In most practical cases, the network is irreducible if every person is a member of the largest strongly connected component and is aperiodic if it is not bipartite.

the posterior distribution,  $\eta_{i,j}^{(1)}, \dots, \eta_{i,j}^{(K)}$ , into predicted probabilities, which I denote  $p_{i,j}^{(1)}, \dots, p_{i,j}^{(K)}$  using the inverse logit transformation:

$$p_{i,j}^{(k)} = \exp\left(\eta_{i,j}^{(k)}\right) / \left(1 + \exp\left(\eta_{i,j}^{(k)}\right)\right) \quad (2.3)$$

for each  $i = 1, \dots, n$ ,  $j = 1, \dots, n$ , and  $k = 1, \dots, K$ . I then create a weight matrix for each of the  $K$  draws from the posterior by setting the  $i^{\text{th}}$ ,  $j^{\text{th}}$  cell of  $W^{(k)}$  to  $p_{i,j}^{(k)}$  and dividing by the row sums to row-normalize  $W^{(k)}$ . I simulate diffusion using the weighted averaging model, meaning  $A^{(k,t+1)} = W^{(k)}A^{(k,t)}$ , where  $A^{(k,t)}$  indicates the value of the attitudes at time  $t$  using the  $k^{\text{th}}$  draw from the posterior distribution. To simulate several iterations of the diffusion process, I hold the weight matrix for a given draw from the posterior distribution,  $W^{(k)}$ , constant, giving  $A^{(k,t+1)} = \left(W^{(k)}\right)^t A^{(1)}$ .<sup>4</sup> This process produces a distribution of attitude values for each person  $i$  at each iteration of the diffusion process  $t$ , for each draw from the posterior distribution  $k$ . Using the different draws from the posterior distribution incorporates the stochastic uncertainty about how close each person is in the underlying space.

Alternatively, setting the elements of  $W$  to a draw from the posterior predictive distribution of the model suggests that, at a single iteration, people are only influenced by people to whom they are directly connected, but that the connections are randomly generated. To incorporate stochastic uncertainty, the connections must change at each iteration of the model. Formally, I take an adjacency matrix drawn from the posterior predictive distribution,  $\tilde{Y}^{(k)}$ , and transform it into a weight matrix  $W^{(k)}$  using the same procedure as the observed adjacency matrix: setting the diagonal equal to 1 and dividing by the sum of the rows. I then use each weight matrix created from a draw from the posterior predictive distribution for one iteration of

---

<sup>4</sup> Note that the index  $k$  is removed from  $A^{(1)}$  because the initial attitude values,  $A^{(1)}$ , are not changed by using different draws  $k$  from the posterior distribution.

the diffusion model. That is,

$$\begin{aligned}
 A^{(2)} &= (W^{(1)}) A^{(1)} \\
 A^{(3)} &= (W^{(2)}W^{(1)}) A^{(1)} \\
 &\vdots \\
 A^{(K)} &= (W^{(K)} \dots W^{(2)}W^{(1)}) A^{(1)}
 \end{aligned}
 \tag{2.4}$$

Thus at each iteration, a new adjacency matrix, representing a new network, is drawn from the posterior predictive distribution, and that draw is used to simulate one iteration of the diffusion process. Stochastic uncertainty about which ties are present is incorporated because at each iteration a different set of ties, drawn from the same set of underlying distances, is used to simulate the diffusion process.

Effectively, both of these methods shrink the diffusion results towards the results from the network predicted by the latent space model. The shrinkage can either happen at the point of the predicted probabilities, as in the first method, or at the point of the predicted networks, as in the second method. The methods also correspond to different assumptions about the diffusion process. The first method assumes that every person can observe every other person’s attitudes, and takes every other person in the network into account when changing his or her mind. . In small groups of less than ten people, this assumption is almost certainly true, but it may be supported in larger groups, such as a school, as well. For example, Fujimoto and Valente (2012) suggest that students in a school mimic structurally equivalent others in the school, rather than people they are directly connected to.

By contrast, the second method extends the existing diffusion model more directly. Using the posterior predictive distribution, meaning the predicted networks drawn from the model predictions, implies that a person only considers his or her immediate network neighbors when he or she changes his or her mind. Unlike in the original weighted averaging model, however, the second method suggests that those

ties are random draws from an underlying probability, and may vary from day to day, or from iteration to iteration. Thus the second model takes the uncertainty in the observed network ties into account by using a different, but similar, set of network ties at every iteration. The following section illustrates how this method is used in practice, using a friendship network from a school.

# 3

## Data

### 3.1 Sample

To illustrate how this method works in practice, I consider a friendship network from a school in the PROMoting School-university-community Partnerships to Enhance Resilience (PROSPER) Peers study (Spoth et al., 2004, 2007). PROSPER Peers is a longitudinal study of adolescents in schools originally designed to test substance use interventions. Researchers identified 28 rural communities — 14 in Iowa, and 14 in Pennsylvania — and interviewed two cohorts of students at each of the schools in those communities. Each cohort was interviewed in the fall of 6th grade, and then was re-interviewed in the spring of 6th grade, 7th grade, and 8th grade, for a total of five wave of in-school interviews. Friendship networks were collected by asking students to name up to two best friends and five other close friends. The names were matched to a class roster to produce a directed friendship network within each grade, where student  $i$  is connected to student  $j$  (denoted  $i \rightarrow j$ ) if student  $i$  listed student  $j$  as a friend. For illustration, I focus on the network constructed from the 6th grade fall survey administration of a single school, which I will call School 212. School 212 is a 6th grade class in Iowa which had 160 students at the time of the

survey, of whom 48.1% were male, 37.8% received free lunch, and 73.8% were white.

### 3.2 Latent space model specification

For simplicity, in this example, I focus on a single model specification, a latent space model with a term for a 2-dimensional Euclidean latent space. I chose this model for two reasons. First, the 2-dimensional latent space model is the simplest, most parsimonious model for the network, which helps clarify the model's presentation. In addition, the 2-dimensional Euclidean latent space model can be readily visualized, allowing visual interpretation of the results as well. Second, I tried additional model specifications, using different numbers of dimensions, as well as matching terms and latent clustering (Handcock et al., 2007) to account for potential in-group biases (Hewstone et al., 2002) within either sociodemographic groups or informal peer groups. Although including different terms in the models may have changed the model fit slightly, they did not change the diffusion results substantially over the simple, 2-dimensional Euclidean latent space model. As such, they have been excluded; future work may examine when additional terms in the model specification produce different diffusion outcomes. The following section shows the results from diffusion simulated using the 2-dimensional, Euclidean latent space model.

# 4

## Results

Figure 4.1 illustrates the diffusion process in School 212 visually. Each column represents one of the diffusion processes: using the observed network, using the posterior predicted probabilities from the latent space model, and using the posterior predictive distribution, meaning the simulated networks, from the latent space model. Each row represents one iteration of the model — the first row represents the initial conditions, the second row represents the attitudes after the weighted averaging has been performed once, the third row represents the attitudes after the weighted averaging has been performed twice, and so on. The coloring of the vertices<sup>1</sup> represents each person’s attitude after each iteration of diffusion process. The initial attitudes are taken from a question asked of the students, “How wrong is it for someone your age to smoke cigarettes?” The response options were coded as 1=Not at all wrong, 2=A little bit wrong, 3=Fairly wrong, 4=Very Wrong; darker colors indicate that a student feels smoking cigarettes is more wrong. To ease comparison, the vertices have the same position, their location in the latent space from the fitted model, in

---

<sup>1</sup> Colors created by Color Brewer (Brewer, 2013). Colors are binned at  $\frac{1}{4}$  point intervals from 1 to 4, i.e., 1.00, 1.25, 1.50, . . . , 3.75, 4.00.

each of the plots.

Figure 4.1 shows the diffusion process for each of the three types of diffusion visually. In the left column, the diffusing bit, attitudes about smoking, spreads along unchanging, observed network ties. Most of the people in the network believe that smoking is very wrong (indicated by dark blue), and that attitude spreads to most of the people who believe smoking is not at all wrong (indicated by light yellow) after the weighted averaging is taken twice. Several people remain unchanged from their initial attitudes, however. These people do not send ties to others in the network, and therefore are not influenced by others in the weighted averaging diffusion model using the observed network after any number of iterations.

The second column shows the weighted averaging process using the predicted probabilities of a tie between each person as a weight. Since each person is effectively connected with each other person, albeit often with very low weights, the network is completely connected in this case. Rather than drawing all of the ties, for clarity, I removed the ties, so that the distances between vertices would be more apparent. Using the predicted probabilities as weights causes the weighted averaging to converge almost entirely after two iterations. Since each person is connected to every other person, each person takes the attitudes of every other person into account during the weighted averaging process, leading to very rapid convergence on a single value for the diffusing attitude.

The third column shows the weighted averaging process using the draws from the predictive distribution. At each step, a new set of ties is drawn from the underlying distances in social space, and then a weighted averaging process is simulated once over that set of ties. This corresponds to an accurate model with unreliably measured ties. In each plot, the set of ties shown are the set of ties that will be used for that instance of the weighted averaging process. As the figure shows, after five iterations, the weighted averaging process has mostly converged on a single consensus

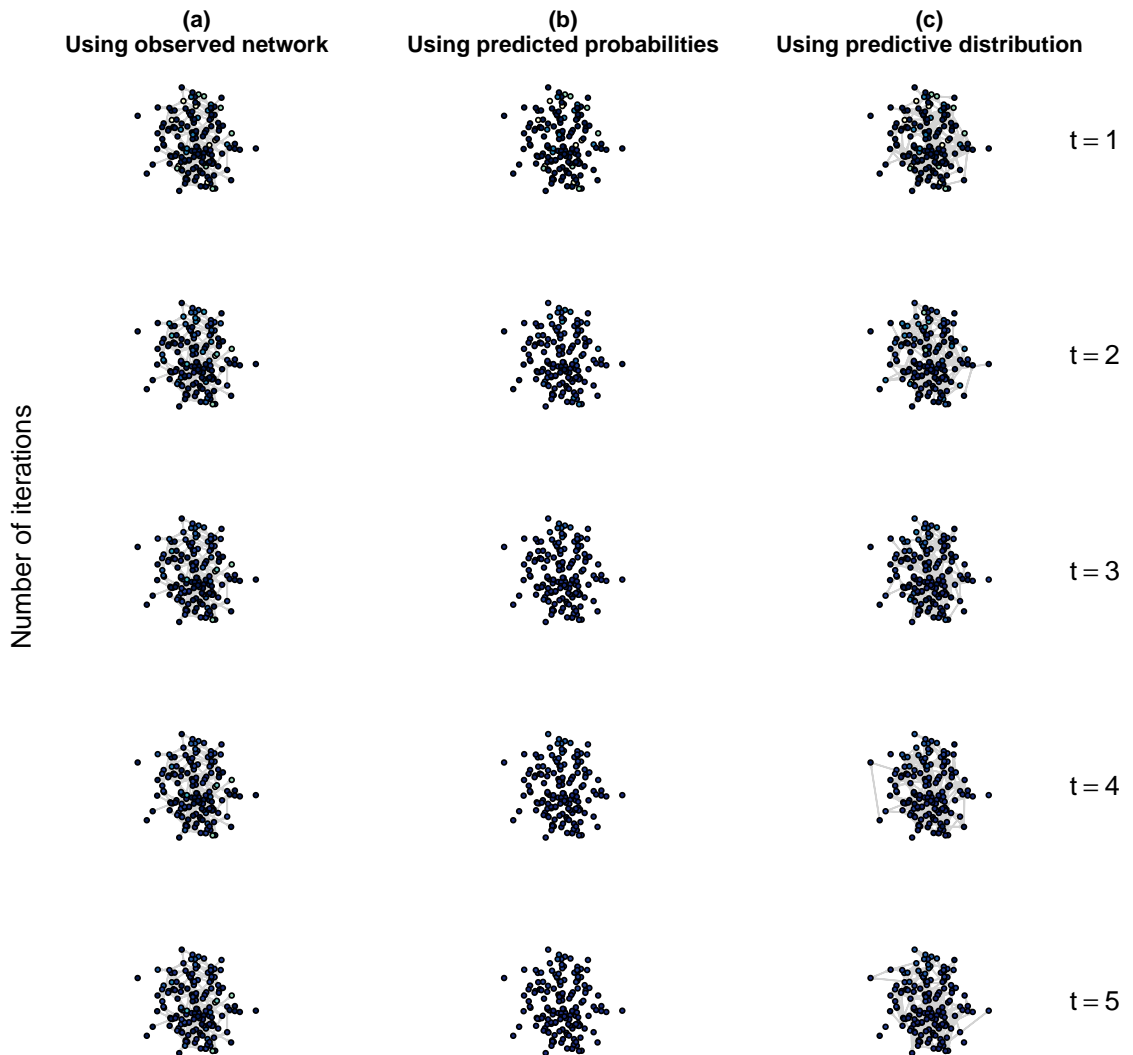


FIGURE 4.1: Visualization of weighted averaging diffusion processes

value. Although a few people who started off believing that smoking was not at all wrong take one or two more iterations to reach the consensus value, all of the people ultimately reach the consensus value. Unlike in the model using the static, observed network, any people who are isolated do not necessarily stay isolated, meaning that they will ultimately contribute to the final consensus in the school. Their contribution is visible both as a slightly lower final value of the consensus attitude, and as a lack of isolates who hold the same attitudes as they did initially.

As Fig. 4.1 shows, the two key differences between the weighted averaging model using the observed network and using the latent space model predictions are the speed with which the weighted averaging converges and the consistency of the final consensus value of the diffusing attitude. Figure 4.2 illustrates the differences in the first aspect, the speed of convergence. The figure shows the change in attitude values between each iteration, calculated as the norm of the difference between the vector of attitudes at time  $t$  and at time  $t + 1$  (logged for scaling), plotted against the number of iterations. The solid line in Fig. 4.2 indicates the changes in the diffusion process run using the unchanging, observed network, while the dashed line shows the changes in the diffusion process using the predicted probabilities, and the dotted line shows the changes in the diffusion process using different draws from the posterior predictive distribution at each step.

Figure 4.2 shows the pattern in rate of convergence as Fig. 4.1. The weighted averaging processes using the observed network and the posterior predictive distribution of the latent space model converge at approximately the same rate. The process using draws from the posterior predictive distribution of the latent space model fluctuates more, because the network changes between each iteration, but ultimately, the models change by approximately the same amount between each iteration. By contrast, the weighted averaging process using the predicted probabilities as weights converges much more quickly. In the predicted probability model, since everyone is connected to everyone else, the weighted averaging process proceeds much more rapidly than when people can only take into account their immediate network neighbors. In both cases the models have mostly converged before ten iterations have been performed.

Both the model based on the observed network and the model based on the simulated networks converge at roughly the same rate, but do they both converge to the same consensus value? Figure 4.3 shows the distribution of attitude values

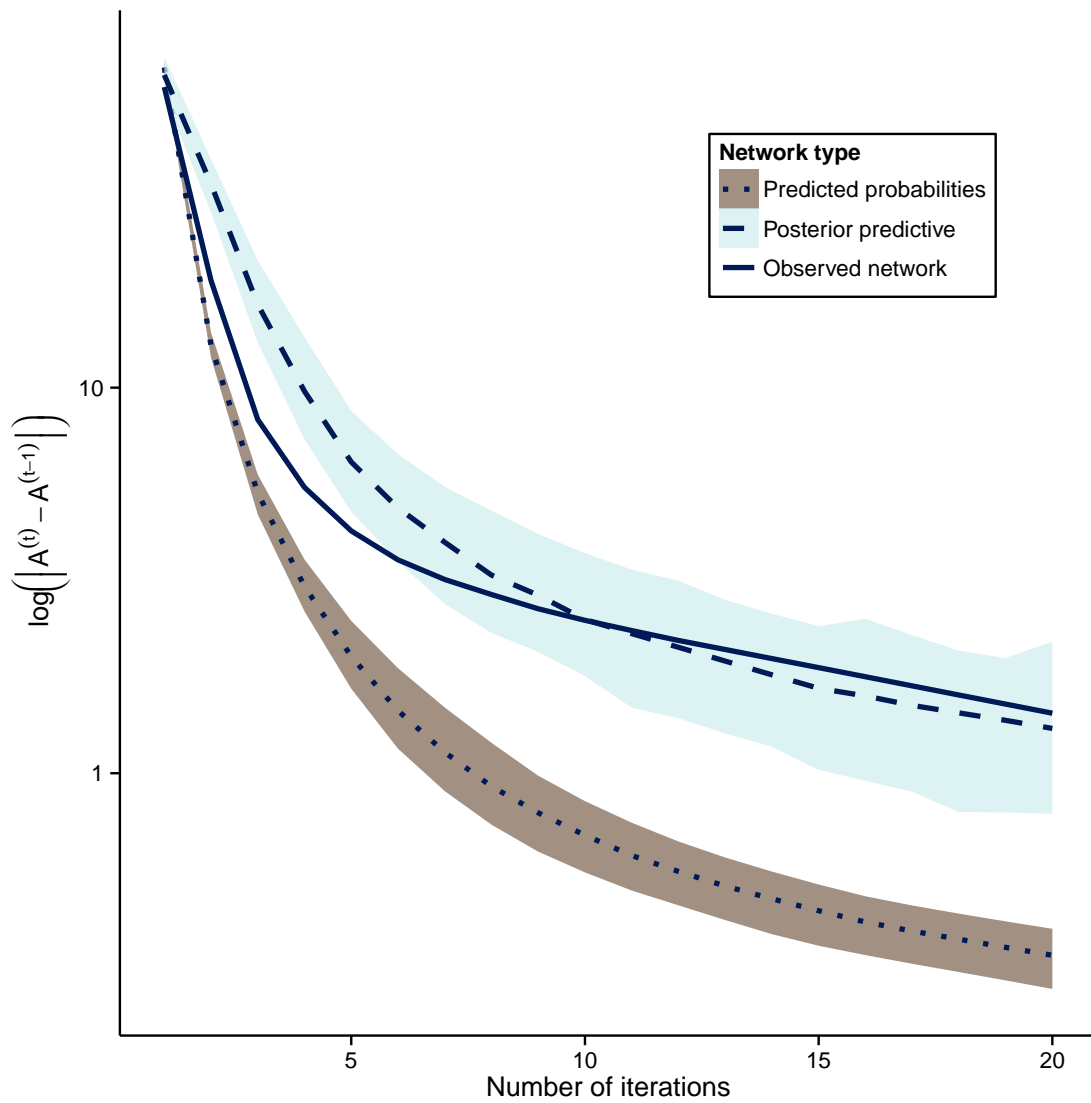


FIGURE 4.2: Change in simulated attitudes after each iteration

at 1, 2, 3, 7, and 11 iterations for each of the models. To avoid confounding the convergence of the model with pre-existing homophily on the an attitude, instead of using values from the "how wrong do you think it is to smoke cigarettes" question as in previous figures, for Fig. 4.3 I randomly assigned each person in the network a simulated attitude value of 1, 2, 3, 4, or 5, and then ran the diffusion process as before.

In each panel of Fig. 4.3, the distribution of attitude values begins as a flat, nearly

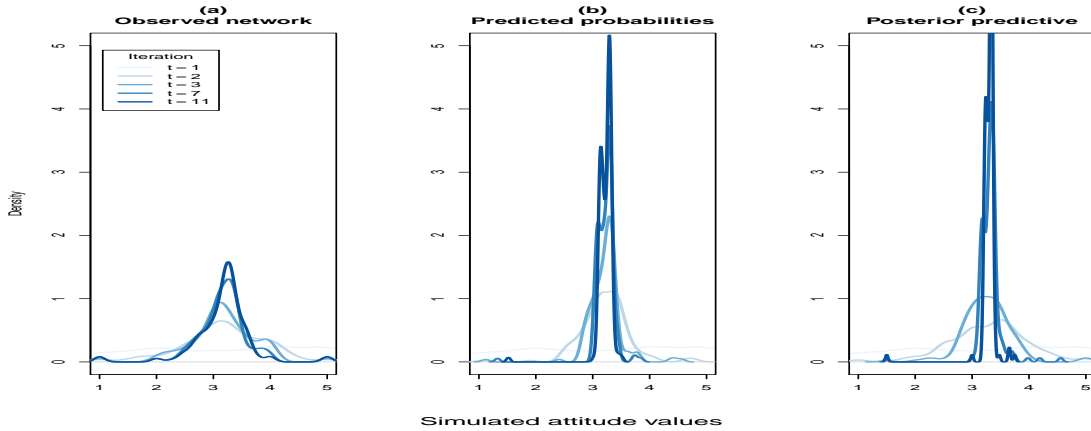


FIGURE 4.3: Distribution of attitude values by simulated iteration

uniform, distribution across the range from 1 to 5. In models (b) and (c), the diffusion processes simulated using the latent space model predictions, the distribution gradually becomes increasingly tightly centered at the mean, approximately 3.5. As in Fig. 4.1 and 4.2, the model using the predicted probabilities, (b), becomes more tightly centered at 3.5 in fewer steps than the model using draws from the posterior predictive distribution, (c). After 2 and 3 steps, the density at 3 is higher for (b) than for (c), and the difference between 7 and 11 steps is still perceptible for (c), while it is not for (b). Unlike the diffusion models simulated from using the latent space model predictions, however, the diffusion model simulated using the observed network, (a), does not converge to a single point as the number of iterations increases. Instead, the model using the observed network develops a distribution of attitudes centered around 3.5, but never achieves the tight distribution at 3.5 that models (b) or (c) do. Unlike in models (b) and (c), people who are disconnected from the network in model (a) never become connected, and therefore maintain their attitudes throughout the simulation. Furthermore, people who do not receive any friendship nominations do not influence anyone, and remain that way throughout the simulation. Model (a) treats the ties as fixed, stable quantities, suggesting that the lack of a tie is always meaningful. As such, model (a) never converges to a single consensus value, while

models (b) and (c) do. Strictly as estimators of the final consensus values, models (b) and (c) are consistent, while model (a) is not.

But which estimator is closer to the truth? Figure 4.4 shows how the estimates of diffusion of attitudes about smoking compare to the attitudes about smoking observed at the second wave of the survey. In Fig. 4.4, the initial attitude values,  $A^{(1)}$ , are given by students' responses to the question "How wrong is it for someone your age to smoke cigarettes?" as in Fig. 4.1 and 4.2. I simulated the diffusion process for 5 iterations using each of the methods, and after each iteration, I calculated the mean squared error (MSE) of the results of the diffusion simulation, and the students responses to the same question at the second survey wave. That is, Fig. 4.4 shows the MSE of the diffusion results and the true values of the diffusing attitude at the next survey administration. Lower values of the MSE indicate that the particular diffusion method is capturing the diffusion process more accurately.

Figure 4.4 shows that diffusion simulated using the predicted probabilities of a tie as weights predicts future attitudes about smoking more accurately than diffusion simulated using either the observed network or the posterior predictive distribution. The MSE for the model using predicted probabilities is approximately 0.1, or 20%, lower than the MSE for the model using either the posterior predictive distribution or the observed network after 1 iteration of the model. Since simulations using the predicted probabilities of a tie allow each person in the network to incorporate information from every other person in the network, while the other two methods do not, Fig. 4.4 suggests that students incorporate information from the other students in the school more broadly, and not just from the people that they are connected to.

Over each iteration, the MSE from the simulation using the posterior predictive distribution declines, until at the 5th iteration, the MSE from the simulation using the posterior predictive distribution and the posterior probabilities of a tie are approximately the same. By contrast, the MSE from the simulation using the observed

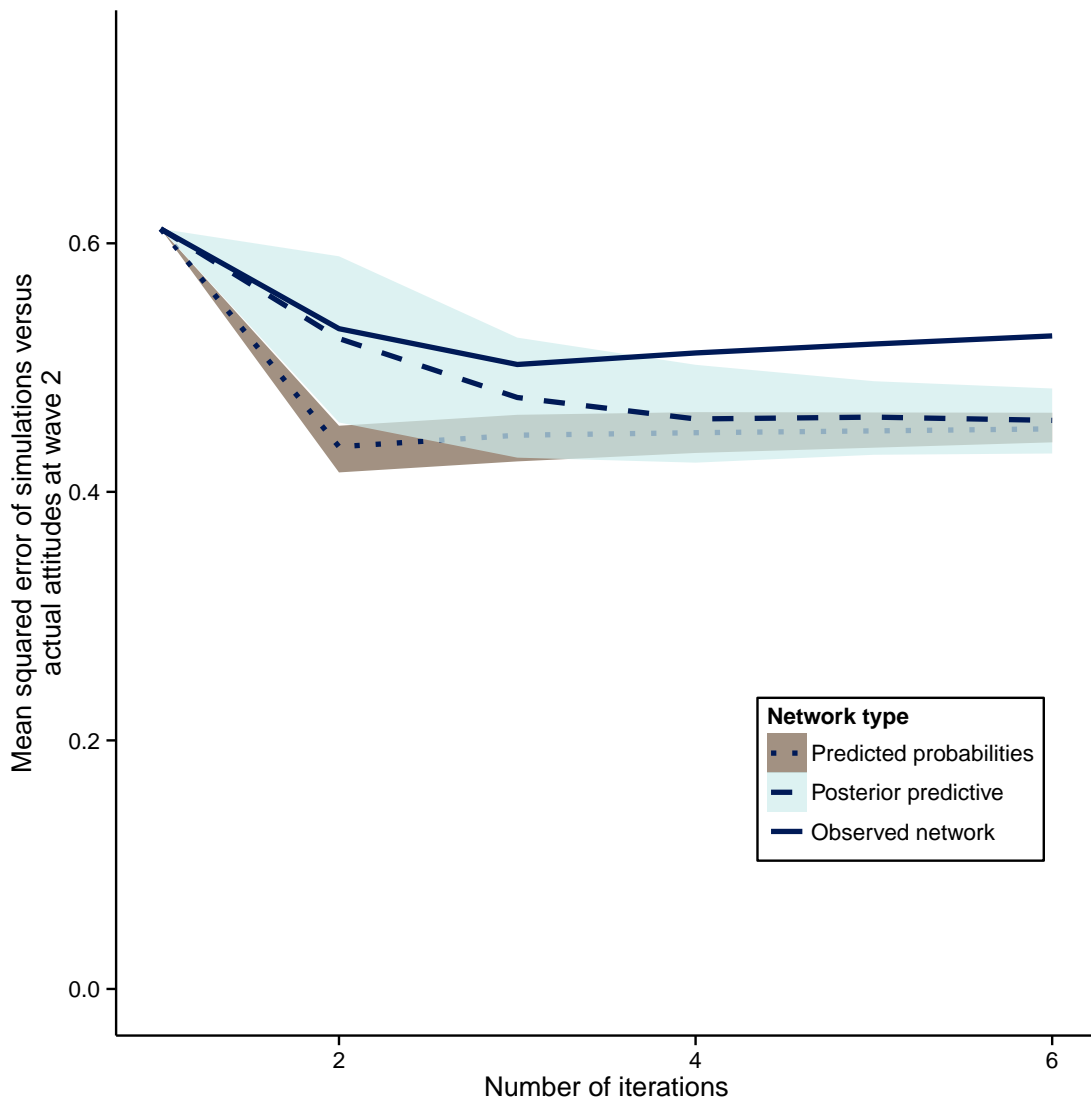


FIGURE 4.4: Mean squared error of predictions

network does not decline appreciably as the number of iterations increases. This follows the results on the convergence of the different models. As Fig. 4.3 showed, the simulations using the predicted probabilities of a tie and the posterior predictive distribution, or the networks drawn from those probabilities, converge to approximately the same consensus value. The approaches differ primarily in the amount of time it takes to reach consensus; since the predicted probabilities average over all the people in the network at each iteration, the simulations using the predicted prob-

abilities converge much faster. Figure 4.4 shows that the consensus value reached by the latent space diffusion model is closer to the true attitude values measured in the second survey administration. The simulations using the predicted probabilities approach that consensus value within 1 or 2 steps, while the simulations using the simulated networks approach the consensus value within 4 or 5 steps. The observed network does not approach a single consensus value, and therefore predicts the final outcome less accurately.

## Discussion

In this article, I have described a latent space diffusion model that integrates two important aspects of how networks are conceptualized: stable connections through which influence and information travel, and indicators of the relative positions of people. The approach described here integrates those perspectives by estimating the relative positions of people to one another using a latent space model, and then simulating the diffusion of an attitude over the latent space. I represented the latent space in two ways: as the predicted probabilities of two people being connected to one another, and as a series of networks randomly drawn from those distances. The two ways refer to a situation where people incorporate information from everyone else in the network, and a situation where people only incorporate information from their immediate network neighbors, but their networks are subject to routine, random churn.

The perspective elaborated here is more consistent, both theoretically and statistically, than using the observed network. Using a latent space model is theoretically more consistent because it represents the measurement error inherent in network observation more accurately. Networks, as we observe them, show continuous, routine,

and rapid churn. Under those conditions, if an attitude or belief spreads through the network, the people in the network are likely influenced by all of the people in their social vicinity, or are influenced by the people they happened to talk to that day. This model addresses both of those possibilities, and shows that the former predicts future attitudes more accurately.

Furthermore, this approach is a consistent estimator of the final consensus that a group would reach over repeated applications of a weighted averaging diffusion process. When only the observed network is used, idiosyncrasies of the ties reported, such as people not receiving any friendship ties, unduly influence the final distribution of attitudes. By contrast, the latent space approach allows random churn in the ties, improving mixing of the people in the model without sacrificing important social distances. Although I focus on the weighted averaging model for simplicity and analytical tractability, this approach could be extended to other diffusion simulation frameworks, such as threshold models and simulation models of culture and structure, in a relatively straightforward manner.

While this model presents a more integrated and theoretically consistent view of network diffusion, it suffers from several limitations. First this model assumes that the observed network is the network that most closely represents the diffusion process. Since the distances in social space are estimated from the observed network, if the observed network does not accurately reflect the diffusion process, then the social space estimated from the observed network will not accurately reflect the diffusion process either. Second, the model assumes that categorical distinctions will be reflected in the observed network proportionally to their salience. If salient categorical distinctions are present, but are not reflected in the network, then the model will not represent in-group biases appropriately. Prior information about the salience of particular categorical distinctions can be incorporated using the prior distributions on model parameters, but I have not attempted that in this study.

Third, the model does not have a way of accounting for structural zeros, or situations where diffusion cannot occur because of separation of the populations. For example, if I had included students from separate schools in the same model, the model would have predicted a positive probability of those students interacting, even if they lived in different states and interaction would have been impossible. I addressed this limitation by fitting each of the models separately; future work could construct a latent space model that accurately reflects structural zeros. Fourth, and finally, the model does not allow the probabilities of interaction to change over time. Instead, changes in the network are seen as different random draws from the same underlying social space, rather than changes in the social space itself. Future work should develop a method that can estimate changes in the social space (cf. Sewell and Chen, 2015), and therefore can estimate how diffusion potential changes over time.

In spite of these limitations, the method proposed in this study represents a significant step forward methodologically. By applying a smoothing function to the observed network, this method also provides an approach to modeling diffusion that is robust to small fluctuations in the observed network, which are a ubiquitous feature of network data collection. Future work may be able to extend this approach to use other models that generate predicted probabilities of network ties, such as the relational events model, to extend the diffusion framework beyond statically observed networks. Networks have long represented two distinct aspects of social structure: the conduits through which information and influence flow, and indicators of positions of positions in social structure, like hierarchy and roles. By constructing a model of diffusion over social space, we can take the first step towards merging these two aspects.

# Bibliography

- Berger, R. L. (1981), “A Necessary and Sufficient Condition for Reaching a Consensus Using DeGroot’s Method,” *Journal of the American Statistical Association*, 76, 415–418.
- Brewer, C. A. (2013), “ColorBrewer: Color Advice for Maps,” .
- Burt, R. S. (1987), “Social Contagion and Innovation: Cohesion versus Structural Equivalence,” *American Journal of Sociology*, 92, 1287–1335.
- Cairns, R. B. and Cairns, B. D. (1994), *Lifelines and Risks: Pathways of Youth in our Time*, Cambridge University Press, Cambridge, OCLC Number: 30892326 Gift of Dorothy Bevan in memory of her husband Professor William Bevan. Includes bibliographical references (p. 288-305) and indexes.
- Centola, D. (2011), “An Experimental Study of Homophily in the Adoption of Health Behavior,” *Science*, 334, 1269–1272.
- Centola, D. and Macy, M. (2007), “Complex Contagions and the Weakness of Long Ties,” *American Journal of Sociology*, 113, 702–734, ArticleType: research-article / Full publication date: November 2007 / Copyright 2007 The University of Chicago Press.
- Chatterjee, S. and Seneta, E. (1977), “Towards Consensus: Some Convergence Theorems on Repeated Averaging,” *Journal of Applied Probability*, 14, 89–97.
- Christakis, N. A. and Fowler, J. H. (2007), “The Spread of Obesity in a Large Social Network Over 32 Years,” *The New England Journal of Medicine*, 357, 370–9.
- Coleman, J. S., Katz, E., and Menzel, H. (1966), *Medical Innovation: A Diffusion Study*, Bobbs-Merrill Co., Indianapolis.
- DeGroot, M. H. (1974), “Reaching a Consensus,” *Journal of the American Statistical Association*, 69, 118–121, ArticleType: research-article / Full publication date: Mar., 1974 / Copyright 1974 American Statistical Association.
- Durrett, R., Gleeson, J. P., Lloyd, A. L., Mucha, P. J., Shi, F., Sivakoff, D., Socolar, J. E. S., and Varghese, C. (2012), “Graph fission in an evolving voter model,” *Proceedings of the National Academy of Sciences*, 109, 3682–3687.

- Eagle, D. E. and Proeschold-Bell, R. J. (2015), “Methodological considerations in the use of name generators and interpreters,” *Social Networks*, 40, 75–83.
- French, J. R. P. J. (1956), “A formal theory of social power,” *Psychological Review*, 63, 181–194.
- Friedkin, N. E. (1998), *A Structural Theory of Social Influence*, Cambridge University Press, Cambridge.
- Friedkin, N. E. and Johnsen, E. C. (2011), *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*, Cambridge University Press,, New York.
- Fujimoto, K. and Valente, T. W. (2012), “Social Network Influences on Adolescent Substance Use: Disentangling Structural Equivalence from Cohesion,” *Social Science & Medicine*, 74, 1952–1960.
- Granovetter, M. (1978), “Threshold Models of Collective Behavior,” *American Journal of Sociology*, 83, 1420–1443, ArticleType: research-article / Full publication date: May, 1978 / Copyright 1978 The University of Chicago Press.
- Granovetter, M. S. (1973), “The Strength of Weak Ties,” *The American Journal of Sociology*, 78, 1360–1380, ArticleType: research-article / Full publication date: May, 1973 / Copyright 1973 The University of Chicago Press.
- Granovetter, M. S. (1995), *Getting a Job: A Study of Contacts and Careers*, University of Chicago Press, Chicago, 2nd edition edn., OCLC Number: 31130973 Includes bibliographical references (p. 241-247) and index.
- Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007), “Model-Based Clustering for Social Networks,” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 170, 301–354.
- Harary, F. (1959), “A criterion for unanimity in French’s theory of social power,” in *Studies in social power*, pp. 168–182, Univer. Michigan, Oxford, England.
- Hewstone, M., Rubin, M., and Willis, H. (2002), “Intergroup bias,” *Annual Review of Psychology*, 53, 575–604.
- Hoff, P. D. (2005), “Bilinear Mixed-Effects Models for Dyadic Data,” *Journal of the American Statistical Association*, 100, 286–295.
- Hoff, P. D. (2009), “Multiplicative latent factor models for description and prediction of social networks,” *Computational and Mathematical Organization Theory*, 15, 261–272.

- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002), “Latent Space Approaches to Social Network Analysis,” *Journal of the American Statistical Association*, 97, 1090–1098.
- Holme, P. and Newman, M. E. J. (2006), “Nonequilibrium phase transition in the coevolution of networks and opinions,” *Physical Review E*, 74, 056108.
- Krivitsky, P. N. and Handcock, M. S. (2008), “Fitting Latent Cluster Models for Networks with latentnet,” *Journal of Statistical Software*, 24, 1–23.
- Martin, J. L. (2002), “Power, Authority, and the Constraint of Belief Systems,” *American Journal of Sociology*, 107, 861–904, ArticleType: research-article / Full publication date: January 2002 / Copyright 2002 The University of Chicago Press.
- McPherson, M. (1983), “An Ecology of Affiliation,” *American Sociological Review*, 48, 519–532, ArticleType: research-article / Full publication date: Aug., 1983 / Copyright 1983 American Sociological Association.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001), “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociology*, 27, 415–444, ArticleType: research-article / Full publication date: 2001 / Copyright 2001 Annual Reviews.
- Moody, J., Brynildsen, W. D., Osgood, D. W., Feinberg, M. E., and Gest, S. (2011), “Popularity Trajectories and Substance Use in Early Adolescence,” *Social Networks*, 33, 101–112.
- Morris, M., Kurth, A. E., Hamilton, D. T., Moody, J., and Wakefield, S. (2009), “Concurrent Partnerships and HIV Prevalence Disparities by Race: Linking Science and Public Health Practice,” *American Journal of Public Health*, 99, 1023–1031.
- Newman, M. E. J. (2006), “Modularity and community structure in networks,” *Proceedings of the National Academy of Sciences*, 103, 8577–8582.
- Paik, A. and Sanchagrin, K. (2013), “Social Isolation in America: An Artifact,” *American Sociological Review*, 78, 339–360.
- Rossman, G., Esparza, N., and Bonacich, P. (2010), “I’d Like to Thank the Academy, Team Spillovers, and Network Centrality,” *American Sociological Review*, 75, 31–51, WOS:000275828900003.
- Schelling, T. C. (1971), “Dynamic Models of Segregation,” *Journal of Mathematical Sociology*, 1, 143–186.
- Schelling, T. C. (1978), *Micromotives and Macrobehavior*, W.W. Norton & Co., New York.

- Sewell, D. K. and Chen, Y. (2015), “Latent Space Models for Dynamic Networks,” *Journal of the American Statistical Association*, pp. 00–00.
- Smith, J. A., McPherson, M., and Smith-Lovin, L. (2014), “Social Distance in the United States Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004,” *American Sociological Review*, 79, 432–456.
- Spoth, R., Greenberg, M., Bierman, K., and Redmond, C. (2004), “PROSPER CommunityUniversity Partnership Model for Public Education Systems: Capacity-Building for Evidence-Based, Competence-Building Prevention,” *Prevention Science*, 5, 31–39.
- Spoth, R., Redmond, C., Shin, C., Greenberg, M., Clair, S., and Feinberg, M. (2007), “Substance-Use Outcomes at 18 Months Past Baseline: The PROSPER CommunityUniversity Partnership Trial,” *American Journal of Preventive Medicine*, 32, 395–402.
- Watts, D. J. (2002), “A Simple Model of Global Cascades on Random Networks,” *Proceedings of the National Academy of Sciences of the United States of America*, 99, 5766–5771.
- White, H. C., Boorman, S. A., and Breiger, R. L. (1976), “Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions,” *The American Journal of Sociology*, 81, 730–780, ArticleType: research-article / Full publication date: Jan., 1976 / Copyright 1976 The University of Chicago Press.