# Mechanisms of Behavioral Contagion: An Approximate Bayesian Approach

Christian C. Luhmann

Department of Psychology Institute for Advanced Computational Science Stony Brook University Stony Brook, NY 11794 Email: christian.luhmann@stonybrook.edu Brian Yang Stony Brook University Stony Brook, NY 11794 Email: wconnection@gmail.com

*Abstract*—Researchers have proposed that contagion processes govern how information and behavior itself spreads through social networks. Empirical evidence for such contagion often makes unjustified, but implicit assumptions about the mechanisms underlying contagion. Here, we present an approximate Bayesian method that uses empirical data to draw inferences about the underlying mechanisms. We provide initial validation of our approach in three simulation experiments, each investigating how a real-world factor (e.g., noise) impacts inferential accuracy.

## 1. Introduction

Social influence has long been known to be a powerful factor shaping people's behavior [1], [2], [3], [4], but researchers have recently suggested that these sorts of mechanisms may give rise to cascades of behavioral change: the notion that behaviors can spread like biological diseases [5], [6]. Behavioral contagion can be advantageous in the case of adaptive behaviors (e.g., vaccination, healthy eating, exercising) but detrimental in the case of maladaptive behaviors (e.g., smoking, needle-sharing, bullying). An accurate understanding of behavioral contagion is critical because it represents a potent tool for policymakers, permitting prediction and control of the social transmission of behavior, both adaptive and maladaptive [7].

Because social influence is inherently unobservable, researchers often make assumptions about the underlying mechanisms and use data to estimate parameters accordingly. These mechanistic assumptions, however, are rarely made explicit or justified. Such work often relates an individual's behavior to the behavior of that individual's peers. For example, studies of contagion often conclude that each additional peer engaging in some behavior (e.g., smoking) incurs some risk for adopting some behavior. Such conclusions are only sensible if the underlying mechanisms are simple, epidemiological-style processes. For example, imagine that individuals only adopt a behavior after a critical mass (of unknown size) of their peers have already adopted the behavior. If this is the case, the probability of adoption "per adopting peer" is nonsensical and highlights the potential difficulty of interpreting peer effects.

## 1.1. Contribution

The current article describes an analytic method for explicitly evaluating assumptions about the mechanisms that underlie large-scale patterns of behavior. We then validate our method by conducting simulation experiments designed to explore a variety of real-world circumstances (e.g., noise). Overall, we find that our method is both successful and reasonably robust. Our method has several specific benefits, including the ability to draw inferences about crosssectional data and the ability to evaluate arbitrary models of behavioral adoption.

## 2. Our Method

We take an approximate Bayesian computation (ABC), or likelihood-free, approach [8], [9]. ABC begins by sampling parameter values from a prior distribution,  $P(\theta)$ , feeding the sampled values into a generative model, m, and generating a synthetic data set,  $\hat{D} = m(\theta)$ . The synthetic data set,  $\hat{D}$ , is then compared to the observed data set, D. If the two are sufficiently similar,  $sim(\hat{D}, D) < \epsilon$ , the corresponding parameter values,  $\theta$ , contribute to the approximation of the posterior,  $P(\theta|D)$ . Because data sets often occupy a high-dimensional space, ABC approaches typically compare summary statistics computed for each data set rather than directly comparing the data sets themselves: e.g.,  $sim\left(\gamma(\hat{D}), \gamma(D)\right)$ . Instead of selecting  $\epsilon$  and sim(), we use a classifier to tailor these components in a data-driven manner.

### 2.1. Models of Behavioral Contagion

Canonical models of both simple and complex contagion were taken from [6]. The simple contagion model, what [6] refers to as the independent cascade model, operates much like standard epidemiological models of disease transmission. Each edge in the network is associated with a transmission probability. When an active node (e.g., A) shares an edge with an inactive node (e.g., B), node A infects node B with the transmission probability,  $p_{transmit}$ , associated with their shared edge.

In the complex contagion model, what [6] refers to as a linear threshold model, each node is associated with an idiosyncratic activation threshold,  $\theta$ . These thresholds sometimes refer to the number of neighbors that must be active (absolute thresholds), but other times refer to the proportion of neighbors that must be active (proportional thresholds). In either case, once the number of active neighbors meets or exceeds a node's threshold, that node becomes active. In the current study, we will explore both absolute and proportional complex contagion processes.

## 2.2. Inference

Given some observed data, D, and a set of models  $m_1, m_2, \ldots m_M$  we wish to estimate  $P(m_i|D) \sim$  $P(D|\theta, m_i) P(\theta|m_i) P(m_i) \forall i \in [1...M]$ . More specifically, we wish to perform model selection on the basis of  $P(m_i|D)$ . One natural way to think about Bayesian model selection is as inference over a hierarchical "metamodel" [10] . In such an approach, specific contagion models (such as those outlined above) are nested within a larger model which has a single, categorical parameter, *i* indexing the constituent models. When i = 1, the metamodel is equivalent to model  $m_1$ . When i = 2, the metamodel is equivalent to model  $m_2$ . We then generate synthetic data sets  $\hat{D}_i \forall i$  (i.e., for each constituent model) and find  $\arg\max sim\left(\gamma(\hat{D}),\gamma(D)\right)$ . We do so by training a classifier to distinguish among the synthetic data sets generated by each generative model (i.e.,  $\gamma(\hat{D}_i) \forall i$ ), and interpreting the output of the trained classifier (i.e., the predicted class) when applied to the observed data as  $P(m_i|D)$ . This approach alleviates the need to explicitly provide components required in the standard ABC approach, such as distance measures, sim(), and tolerances,  $\epsilon$ . Instead, we allow the classifier to optimize both the similarity function and the decision

#### 2.3. Networks

boundary.

To ensure results are applicable to a variety of realworld social network structures, we will simulate the various contagion processes on both scale-free [11], [12] and smallworld networks [13]. These particular types of networks reflect important characteristics of real-world social networks (e.g., degree distribution, average path lengths, and clustering). All networks were constructed so as to have an average degree of four and to have 1,024 nodes (except in the first experiment which investigated the influence of network size).

## 2.4. Simulation Details

For each individual simulation, a fresh network was generated. A small number of nodes was then made active, seeding the subsequent contagion processes. Seeding involved activating a single, randomly selected node, as well as all of that node's neighbors. After seeding, each of the three contagion processes under consideration (i.e., simple, absolute complex, and proportional complex) were initiated, each contagion process using the same exact network and the same exact initial conditions (i.e., identical seed nodes). All contagion processes were halted once at least 50% of the network's nodes were active. This stopping condition prevents trivial features like global adoption rate from being used to infer the underlying contagion type. We ultimately generated 1000 data sets for each combination of contagion type and network.

#### 2.5. Summary Statistics

Like traditional ABC methods, it is critical that the statistics used to summarize the data sets (i.e.,  $\gamma$  above) be diagnostic of the parameter(s) of interest (here, the indicator variable representing which model generated the observed data).

- 1) Degree of each node
- 2) Number of each node's neighbors that are active
- 3) Proportion of each node's neighbors that are active
- 4) Number of triangles each node is involved in
- 5) Betweenness centrality of each node
- 6) Closeness centrality of each node
- 7) Harmonic centrality of each node
- 8) Eigenvector centrality of each node
- 9) Katz centrality of each node
- 10) Number of cliques each node is involved in
- 11) Density of the subgraph active/inactive nodes

In total, we used 22 separate statistics to summarize each data set: the mean values of 1-11 for **active** nodes and the mean values of 1-11 for **inactive** nodes. The summary statistics were then standardized and used as features in the classification scheme described below.

Note that many of these features will, in general, be strongly correlated. Though not entirely desirable, there are several reasons to not be overly concerned with such correlations. First, our intention at this stage was to be expository rather than to optimize performance. Second, we evaluate our method by comparing out-of-sample predictions to ground truth parameter estimates, meaning that redundant predictors are not a particular threat. Finally, we use regularization in our classifier (see below), which should both help to alleviate overfitting and help to minimize other undesirable consequences of dependent features. We expect that future work will find a variety of modifications that provide substantial improvements to computational efficiency, including a more thoughtfully curated set of features.

## 2.6. Classification Scheme

To distinguish among the data sets generated by the different transmission models, we employed regularized multinomial logistic regression. We specifically used a oneagainst-all classification scheme, with each regression model tasked with distinguishing one of the three classes from the remaining two classes. More sophisticated classification methods (e.g., support vector machines) were investigated, but they did not improve performance and do not offer disciplinary analysts the straightforward interpretability of logistic regression. Classification performance was evaluated using a stratified 5-fold cross-validation scheme.

#### 2.7. Priors

Of all the model parameters, we are primarily interested in a single parameter: the model index parameter i. In the current illustration, we have no particular prior beliefs and thus sample i uniformly. Thus, we stratified data, ensuring that the training and testing sets each included an equal number of data sets for each of the three models under consideration.

The simple contagion model requires specifying  $p_{infect}$ for each edge in the network. The complex contagion model (both absolute and proportional) requires specifying an activation threshold for each node in the network. Here again, we are agnostic regarding parameters values and specified  $p_{infect} \sim U(.5,1)$  and  $\theta \sim U(0,0.5\mu_{deg})$ , where  $\mu_{deg}$ was the average degree of nodes in the network under study. We would have opted for even wider ranges (e.g.,  $p_{infect} \sim U(0,1)$  or  $\theta \sim U(0,1\mu_{deg})$ ), but such ranges yield degenerate data sets with high probability (e.g., diffusions cannot be propagated by nodes with  $p_{infect} = 0$  or  $\theta = 0$ ). Such degeneracy dramatically increases the computational demands (many simulations will yield data sets that are immediately thrown out). Thus, these distributions represent a compromise between our true agnostisism and practical concerns.

## 3. Network Size

We begin our investigation into practical factors by exploring how the size of the network might constrain classification performance. In actual practice, researchers may not be able to choose how large their data set is for a variety of practical reasons. However, by providing information about the relationship between network size and performance, researchers using our method should be able to determine, a priori, whether the particular network(s) they wish to make inferences about are problematically small.

We generated both scale-free and small-world networks of various sizes (25 - 5625 nodes). For each combination of network type and size, we generated 3,000 separate data sets, 1,000 for each of the three types of contagion (e.g., simple, absolute complex, proportional complex). Classification performance was assessed using five-fold cross-validation as described above. In each of the five folds, 2,400 of the data sets were used for training and 600 were used to test predictive performance. The data was again stratified such that the subset used for testing contained exactly 200 data sets generated by each of the three types of contagion.



Figure 1. Results from our manipulation of network size. We used two types of networks and nine sizes. The dashed horizontal line represents chance-level performance. Note that the x-axis is logarithmic.

### 3.1. Results

Results (Figure 1) illustrate that classification performance was well above chance for all network sizes, was greater than 90% for networks with at least 256 nodes, and steadily increased with network size. In addition, performance was uniformly better in in the scale-free networks than in the small-world networks. This difference may be due to the heterogeneity of scale-free networks [14], though future work is needed to explore this possibility. The current results suggest that our method can provide accurate insights for networks of even modest size (e.g., those collected using more laborious sociometric methods).

## 4. Adoption

We next investigated how classification performance was related to global levels of adoption (sometimes referred to as saturation). In the simulations described above, all contagion processes were halted once the proportion of active nodes surpassed 50%. In the real world, however, behaviors of interest will vary widely in how pervasive they are: obesity is relatively common whereas use of a new recreational narcotic may be relatively rare. In a real-world application of our method, we envision researchers constraining their simulations such that they ultimately yield levels of adoption that mirror those seen in their empirically-observed data set. For this reason, level of adoption, like network size, is unlikely to be a parameter researchers tweak to improve performance. However, the results of this investigation will speak to how effective our method may be in a given application.

We generated both scale-free and small-world networks, each with 1,024 nodes. As the previous simulation's results illustrate, this network size is large enough to support successful inference, but small enough to be relevant in a wide variety of applications. We simulated the three different types of contagion (e.g., simple, absolute complex, proportional complex), halting these processes at different



Figure 2. Results from our manipulation of adoption level. The dashed horizontal line represents chance-level performance. Note that the x-axis is logarithmic.

points. For each combination of network type and saturation threshold, we generated 3,000 separate data sets, 1,000 for each of the three types contagion. Classification performance was again assessed using stratified five-fold cross-validation.

#### 4.1. Results

Classification performance steadily increased with the proportion of active nodes (Figure 2). Performance ranged from approximately 55% at the lowest levels of adoption (0.5%) to approximately 95% correct once approximately 33% of nodes were active. Accuracy of approximately 75% was achieved when only 5% of nodes were active. Performance did not seem to depend strongly on the network types. These results suggest that levels of adoption strongly impact classification accuracy, but that information can be extracted from even modest levels of adoption.

## 5. Noise

The simulations, summary statistics, and classification scheme were identical to the investigations reported above. All simulations were halted once 50% of nodes became active. After the simulations were complete, but before calculating the summary statistics, noise was added to each data set. Specifically, we flipped a biased coin for each node in the network. With probability  $p_{noise}$ , the state of the node was flipped (i.e., an active nodes modified to be inactive or an inactive node modified to be active). We explored a range of values  $p_{noise}$ , providing insight into how robust our method is to contamination (e.g., measurement error).

#### 5.1. Results

Classification performance steadily decreased as the amount of added noise increased (Figure 3). At low levels of noise, classification accuracy was at levels reported above, steadily decreasing to chance levels of performance (i.e., 33%) once 50% of nodes were in the "incorrect" state.



Figure 3. Results from our manipulation of noise. The dashed horizontal line represents chance-level performance.

As with the network size experiment (but unlike adoption rate), performance was uniformly superior when making inferences about scale-free networks.

## 6. Conclusion

We have presented an Approximate Bayesian method for making inferences about the mechanisms of behavioral transmission that underlie empirically observed data sets. Our method is based on traditional approximate Bayesian methods [8], [9], [15], but uses classification methods to automatically make decisions about certain aspects of the procedure (e.g., data set similarity).

To evaluate our method, we have conducted several simulation studies. Specifically, we investigated how various practical constraints might limit the performance of our method. These constraints included the size of the network, the level of adoption (saturation), and the presence of noise. Each of these factors qualified the inferential accuracy of our method. For example, larger networks yielded better performance (nearly perfect in our largest networks), though performance for small networks was still above chance. Networks in which many individuals have adopted the target behavior were also yielded more accurate inference. Finally, contaminating the data sets with random noise hurt inferential accuracy (as expected), but our methods were robust to substantial adulteration. These results demonstrate how real-world factors impact inferential accuracy and highlight how researchers would evaluate accuracy in their particular application before ever applying it to their own data.

Our method can be applied in cases where analysts possess a) social network information and b) node-level adoption information bot of which are cross-sectional. Analysts would then implement candidate models of social contagion based on knowledge of the application domain. These candidate models would be used to generate simulate contagion on the empirically observed network. This simulated data would be used to train and validate a classifier. If performance was satisfactory, the trained classifier would be applied to the empirically observed network, providing inference about the most likely generative mechanisms.

The method described here is presented as a contrast to the conventional practice of analyzing social network data with often implicit assumptions about the mechanisms of social contagion. Our method may seem to be exclusively datadriven, but the description of how to applying our method should make it clear that our method can only adjudicate among provided, candidate mechanisms. Analysts cannot simply provide data and ask, "what sorts of mechanism generated this data?" Our method certainly uses data to answer important analytic questions, but those questions do need to be formulated in sensible ways.

Our method has several obvious benefits. First, our method permits unsatisfactory classification accuracy to be detected before it is applied to empirical data. Second, our method provides information about how details of the generative model(s) shape network-level patterns. For example, social scientists can inspect the coefficients in the regression model to reveal which network-level summary statistics are diagnostic of any particular contagion mechanism. Finally, our method is thoroughly Bayesian and thus comprises all the benefits (and drawbacks) of standard Bayesian approaches. Priors are reflected in the parameter sampling scheme. It is assumed that domain experts could supply relevant guidance regarding prior beliefs. Of course, uninformative prior distributions can be used instead (as was done in the current experiments).

### Acknowledgments

This material is based on work supported by National Science Foundation Grant #1456928. The authors would like to thank Jason Jones, Kelli Johnson, Petar Djurić, Sean Clousten, and Jie Gao for helpful comments on previous versions of this work. The authors also would like to thank Stony Brook Research Computing and Cyberinfrastructure, and the Institute for Advanced Computational Science at Stony Brook University for access to the high-performance SeaWulf computing system, which was made possible by a \$1.4M National Science Foundation grant (#1531492).

### References

- S. E. Asch and H. Guetzkow, "Effects of group pressure upon the modification and distortion of judgments," *Groups, leadership, and men*, pp. 222–236, 1951.
- [2] S. Milgram, L. Bickman, and L. Berkowitz, "Note on the drawing power of crowds of different size." *Journal of Personality and Social Psychology*, vol. 13, no. 2, p. 79, 1969.
- [3] R. B. Cialdini and N. J. Goldstein, "Social influence: Compliance and conformity," *Annual Review of Psychology*, vol. 55, pp. 591–621, 2004.
- [4] M. Sherif, The Psychology of Social Norms. Harper, 1936.
- [5] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 199–208.

- [6] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM* SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003, pp. 137–146.
- [7] T. W. Valente, "Network interventions," *Science*, vol. 337, no. 6090, pp. 49–53, 2012.
- [8] S. A. Sisson, Y. Fan, and M. Beaumont, *Handbook of Approximate Bayesian Computation*. Chapman and Hall/CRC, 2018.
- [9] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate bayesian computation in population genetics," *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002.
- [10] L. Wasserman *et al.*, "Bayesian model selection and model averaging," *Journal of mathematical psychology*, vol. 44, no. 1, pp. 92–107, 2000.
- [11] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [12] M. E. Newman, "Clustering and preferential attachment in growing networks," *Physical Review E*, vol. 64, no. 2, p. 025102, 2001.
- [13] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'smallworld'networks," *Nature*, vol. 393, no. 6684, p. 440, 1998.
- [14] R. Jacob, K. Harikrishnan, R. Misra, and G. Ambika, "Measure for degree heterogeneity in complex networks and its application to recurrence network analysis," *Royal Society open science*, vol. 4, no. 1, p. 160757, 2017.
- [15] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder, "Approximate bayesian computational methods," *Statistics and Computing*, vol. 22, no. 6, pp. 1167–1180, 2012.