SEGREGATION IN SOCIAL NETWORKS: MARKOV BRIDGE MODELS AND ESTIMATION

Vikram Krishnamurthy, Rui Luo, Buddhika Nettasinghe

School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA

ABSTRACT

This paper deals with the modeling and estimation of the sociological phenomena called segregation in social networks. Specifically, we present a novel community-based graph model that represents segregation as a Markov bridge process. A Markov bridge is a one-dimensional Markov random field that facilitates modeling the formation and disassociation of communities at deterministic times which is important in social networks with known timed events. Based on the proposed model, we provide Bayesian filtering algorithms for recursively estimating the level of segregation using noisy samples obtained from the graph. Numerical results indicate that the proposed filtering algorithm outperforms the conventional hidden Markov modeling in terms of the mean-squared error. The proposed filtering method is useful in computational social science where data-driven estimation of the level of segregation from noisy data is required.

Index Terms—Markov bridge, Bayesian filtering, social networks, Schelling’s model, segregation

1. INTRODUCTION

This paper studies statistical signal processing methods for modeling and estimating the sociological phenomena of segregation and integration in social networks. We explore a social network containing multiple communities where the weight of the inter-community edges (that indicate the strength of connectivity between them) evolves randomly over time as a Markov bridge. As discussed in Section 2, the Markov bridge dynamics for the weights in a graph yields a useful model for explaining phenomena in marketing and computational sociology where a social network segregates into communities at a specific time instant and integrates into a single community at another time instant. We propose a Bayesian filtering algorithm to estimate the state of the social network in between these time instants. In this context, the main results of this paper are as follows:

1. We present a dynamic network formation model that capture the dynamics of how a social network segregates into communities (and then integrates back again). The key idea behind our model is to represent the strength of the ties between communities (in terms of a graph clustering metric) as a Markov bridge process.

2. Based on the above model, we propose an inhomogeneous Bayesian filter (called Hidden Markov bridge filter) for recursively estimating the state of the graph clustering metric. The Hidden Markov bridge filter uses only a few (compared to graph size) noisy samples from the social network at each time instant.

3. We numerically compare the performance the proposed Hidden Markov bridge (HMB) filter with conventional Hidden Markov Model (HMM) filter in terms of mean-squared error. Our results show that the proposed method outperforms the traditional HMM filter. This shows that the Bayesian filter yields useful real-time information about the dis-association and association of communities in a network.

Motivation and Context: The temporal dynamics of social networks give rise to states where the network is segregated into multiple communities at certain time instants and integrate back into a single community at other time instants.
low, we discuss two hypothetical examples of such situations which motivate the model that we propose in this paper.

**Example 1. Schelling’s Segregation Model:** The segregation model developed by Schelling [1] is set in an $N \times N$ grid. Agents are split into two groups and occupy the spaces of the grid. Agents desire a fraction $B_0$ of their neighborhood to be from the same group. The threshold value of $B_0$ was approximately $1/4$ to achieve a segregated configuration. The physical grid space can be extended to a social network as a grid graph. Social events such as elections will increase $B_0$ and lead to polarization and weaker inter-community connectivity.

**Example 2. Social Media Marketing Model:** Consider a social media marketing scenario (e.g. Facebook Business page) where a company is connected with customers. Customers are classified into fans and utilitarian customers [2]. As depicted in Fig.1, while fans (bottom left vertices) have a stable connection strength (i.e. fixed edge weight) with the company (center vertex), utilitarian customers (top right vertices) have time-varying connection strength (i.e. time varying edge weights) with the company due to reasons such as sales events. The variable connection leads to segregation and integration of the company-customer social network.

**Related Work:** Several works have studied social segregation. Examples include: [3] that aims to predict the emergence of segregation, [4] that analyzes its effects on opinion polarization. Other works have studied on the effects of segregation from marketing perspective[5, 6, 2, 7]. In the context of such prior work, the aim of this paper is to provide a tractable model for social segregation (in a company consumer network) and propose a Bayesian estimation method when the dynamics are observed only via noisy samples of the network.

2. **MARKOV BRIDGE MODEL FOR DYNAMIC SOCIAL NETWORKS**

This section presents a stochastic model to represent the evolution of a social network whose state is fixed at the beginning and at the end. The two fixed states correspond to a segregated social network (with multiple communities) and a social network which has a single community (i.e. an integrated network). Thus, the model presented in this section is a simple, intuitive representation of the process of social network segregation. Further, as we show later in Section 3, the proposed model is easily amenable to Bayesian statistical inference, making it useful in data-driven contexts in computational social science.

2.1. **Social Network as a Time Varying Graph Model**

This subsection explains the time-varying graph model using a company-customer social network consisting of a company and two types of nodes as the setting.

The social network at discrete time instant $t$ is modeled by an undirected, weighted graph $G(t) = (V, E, w(t))$, with $|V|$ number of agents, $|E|$ number of undirected edges representing their connectivity, and $w(t) : E(t) \rightarrow \mathbb{R}^+$ representing the weights of the edges (i.e. the strengths of each connection). Let $|V| = n$ be the number of vertices in the network, $M = \{v_1, \ldots, v_m\} \subset V$ be the set of utilitarian customers who are all connected with each other (i.e., form a complete subgraph), $v_{m+1} \in V$ be the company, and $N = \{v_{m+2}, \ldots, v_n\} \subset V$ be the set of other customers (fans) that form another complete subgraph. Then, edge weight function $w_{ij}(t)$ between $v_i$ and $v_j$ where $(v_i, v_j) \in E$ is as follows:

$$w_{ij}(t) = \begin{cases} W_{ij}(t) & \text{if } v_i = v_{m+1} \& v_j \in M \\
1 & \text{if } v_i = v_{m+1} \& v_j \in N \\
1 & \text{if } (v_i, v_j) \in M \& v_i \neq v_j \\
1 & \text{if } (v_i, v_j) \in N \& v_i \neq v_j \\
\end{cases}$$

where, $W_{ij}(t), t = 1, 2, \ldots$ is the Markov bridge process that we define in Section 2.2.

Note that (1) classifies the edge weights into three groups: between company and utilitarian customers, between company and fans, and between two customers of the same type. Note that there is no edge between two customers of different types. The edge weights between company and utilitarian customers is subject to sales events and therefore described as a time-evolving random process $W(t)$ (specified in Section 2.2). Other weights are simply set to be one. The simplification is reasonable because fans would be indifferent about sales events and have a more stable relationship with the company. Further, we also assume that the customers of the same type are all connected with each other motivated by the concept of homophily [8].

2.2. **Markov Bridge Model of Edge Weights**

We now propose a Markov bridge model for the evolution of the weight $W_{ij}(t)$ in the graph. Recall [9] a Markov bridge (MB) is a one-dimensional Markov random field. It is clamped at the beginning and end time point and evolves in between with a three point transition probability $p(W_{ij}(t)|W_{ij}(t + 1), W_{ij}(t - 1))$. A Markov bridge model for $W_{ij}(t)$ facilitates modeling a community that separates and then reintegrates with another community in a network. Unlike a Markov chain which enters a state at a geometrically distributed time, a Markov bridge enters a state at a fixed deterministic time [10].

We consider $(2T - 1)$ time steps as the period between consecutive sales events. The edge weight $W_{ij}(t)$ between company and utilitarian customers reaches maximum at time $1$ and time $2T - 1$ when sales event happens, and decreases to minimum at time $T$ in the middle of two consecutive sales
events. The process can be described as a two consecutive MBs as we explain next.

The Markov process $W_{ij}(t)$, $t = 1, 2 \cdots, 2T - 1$ takes values in some finite set $S = \{0, \frac{1}{T-1}, \cdots, \frac{T-2}{T-1}, 1\}$ which is an arithmetic sequence with $T$ elements. Two MBs (for time instants $1$ to $T$ and $T$ to $2T - 1$) are formed by conditioning the Markov process $W_{ij}(t)$ taking fixed states at time instants $1$, $T$ and $2T - 1$. The transition matrix for the reference Markov process is chosen to be a row-normalized Toeplitz matrix such that transitions to neighboring regions are more likely. Let the entries of the matrix be $P_{a,b} = P\{W_{ij}(t+1) = b|W_{ij}(t) = a\}$ for all $W_{ij}$ in (1). Then, for the first MB from $t = 1$ to $t = T$, the transition probabilities are obtained by applying Bayes rule as follows [9]:

$$B_{a,b}^c(t) = P\{W_{ij}(t) = a, W_{ij}(T) = c\} = \frac{P_{a,b}(PT^{-1})_{b,c}}{(PT^{-1})_{a,c}}$$

(2)

for $t = 0, \cdots, T - 2$. The MB has fixed initial state and final state, i.e. $W_{ij}(1) = 1$, $W_{ij}(T) = 0$. The second MB can be constructed similarly. Thus, the dynamics of edge weights is specified by two MBs with transition probability matrices given by (2) and fixed initial state.

### 2.3. Graph Clustering Metrics

The aim of this subsection is to discuss the graph metric called graph conductance that we use to express segregation. Graph conductance is a measurement of the level of clustering in a graph and is explained below.

The conductance of a cut $(S, \overline{S})$ in a graph is defined as:

$$\phi(S) = \frac{\sum_{i \in S} \sum_{j \notin S} w_{ij}}{\min \{a(S), a(V \setminus S)\}}, \ S \subset V$$

(3)

where $a(S) = \sum_{i \in S} \sum_{j \in V} w_{ij}$ is the sum of the weights of all edges with at least one endpoint in $S$. Then, the graph conductance is,

$$\phi(G) = \min_{S \subset V} \phi(S).$$

(4)

Graph conductance is also related to the algebraic connectivity which is the second-smallest eigenvalue of the Laplacian matrix of $G$. Algebraic connectivity is used in many results in spectral graph theory such as Cheeger’s inequality [11]. Fig. 2 shows the variation of both graph conductance and algebraic connectivity follow a similar dynamics. This implies that an estimate of the graph conductance also serves as a proxy for the algebraic connectivity under our model.

### 3. BAYESIAN ESTIMATION OF GRAPH METRICS

Section 2 presented a Markov bridge based model for a social network segregation. A natural question is: assuming the Markov bridge model of Section 2, how can one estimate the level of segregation in a data driven manner? An answer to this question is useful in computational social science and network science that deal with large scale, partially observable (via noisy samples) social networks. As a solution, we propose a Bayesian filtering method based on the proposed segregation model.

#### 3.1. Measuring Conductance via Sampled Edges

This subsection discusses our criteria for obtaining a noisy estimate of the conductance of the underlying dynamic graph (explained in Section 2) using a sampled subgraph.

We assume that $\gamma N$ of the total $N$ edges are uniformly sampled and observed at each time $t$ (random sampling of edges has been used widely in literature in statistical estimation tasks e.g. [12, 13, 14]). $\gamma$ is a fixed ratio in $(0, 1]$. The observed graph conductance $\phi(G(t))$ is computed from the partially sampled graph $G(t)$ at time $t$. Graph conductance is a static function of edge weights,

$$\phi(G(t)) = f(w_{11}(t), \cdots, w_{NN}(t))$$

(5)

And the observed graph conductance is the same function of sampled edge weights,

$$\hat{\phi}(G(t)) = f(w_{11}(t), \cdots, w_{NN}(t))$$

(6)

where $i_1, \cdots, i_N$ are sampled from $1, \cdots, N$ with equal probabilities. From (5), it follows straight forwardly that graph conductance as a static function of the edge weights, follows the same MB dynamics. For the rest of the paper, we denote $\phi(G(t))$ and $\phi(G(t))$ as $\phi(t)$ and $\hat{\phi}(t)$ respectively.

To estimate the observation probabilities $p(\phi(t) = j)$, we use a Monte Carlo simulation to obtain sample trajectories of the $(2T - 1)$-step graph evolution and compute the empirical CDF of noise of the conductance computed from the partial observation i.e. the CDF of the difference between the estimated conductance $\gamma \phi(t)$ and the true conductance $\phi(t)$. Fig. 3 shows that the observation noisy is approximately (in the sense of Kolmogorov-Smirnov test) a Gaussian distribution i.e.

$$p(\hat{\phi}(t)|\phi(t) = i) \sim \mathcal{N}(\gamma \phi(t) - \phi(t)|\mu(t), (\sigma(t))^2)$$

(7)
Fig. 3. The empirical CDF of the sampling noise for graph conductance and the CDF of a Gaussian distribution at four time instants during a simulation of $T = 20$ steps. The $p$-value of KS test indicates that the observation noise can be approximated by Gaussian noise.

3.2. Hidden Markov Bridge Filter

In this subsection, we exploit the Gaussian approximation of measurement noise (obtained in Section 3.1) for recursively tracking the state of the conductance using a Hidden Markov bridge (HMB) filter. HMB filter is a generalization of the time-homogeneous HMM filter [15] and have been widely used in signal processing methods for target tracking [16, 17, 18].

Suppose that the MB process $\Phi = \{\phi^{(1)}, \ldots, \phi^{(t)}\}$ is observed via the observation process $\tilde{\Phi} = \{\tilde{\phi}^{(1)}, \ldots, \tilde{\phi}^{(t)}\}$. Assume that the observation at time $t$ given the state $\phi^{(t)}$ is conditionally independent of $\phi^{(\tau)}$ and $\tilde{\phi}^{(\tau)}$, $\tau \neq t$. This conditional independence implies that

$$P(\phi^{(1)}, \ldots, \phi^{(t)}) = \prod_{k=0}^{t} P(\phi^{(k)}) = \prod_{k=0}^{t} P(\tilde{\phi}^{(k)}) \quad (8)$$

The process $\tilde{\Phi}$ is called a Hidden Markov bridge (HMB) because the property (8) is analogous to the assumption made for Hidden Markov Chains (HMCs). Consider the HMB $\Phi$ with state $\Phi$, known MB transition probability (2), and precomputed observation probability (7). The filtered posterior probability can be evaluated recursively via Bayes’ rule

$$q_{j}(t+1) = \frac{p(\tilde{\phi}^{t+1}|\phi^{t+1} = j) \sum_{i=1}^{\Phi} B_{i,j}(t) q_{i}(t)}{\sum_{i=1}^{\Phi} p(\tilde{\phi}^{t+1}|\phi^{t+1} = l) \sum_{i=1}^{\Phi} B_{i,j}(t) q_{i}(t)} \quad (9)$$

as shown in [19].

4. NUMERICAL RESULTS

In this section, we numerically illustrate that the proposed HMB filter (Section 3) outperforms (in terms of mean-squared error) the widely used HMM filter for estimating the level of segregation. This highlights how the proposed model and filtering method can be useful in estimating the level of segregation with a better accuracy compared to the baseline method of HMM filtering.

Simulation setup: We consider a company-customer network of 10 utilitarian customers, 20 fans, and 1 company as discussed in Section 2 for $2T - 1$ time steps. The state space of the weight of each edge between utilitarian customers and company is an arithmetic sequence $\left[1, \frac{T-1}{T}, \cdots, \frac{1}{T}, 0\right]$. The weight evolves according to a transition matrix which is a Toeplitz matrix. Each descending diagonal from left to right is constant: $\left[(\frac{1}{4})^{T-1}, \cdots, 1, \cdots, (\frac{1}{4})^{T-1}\right]$. Each row vector of the Toeplitz matrix is normalized so that the row elements add up to 1. We then implement the HMB filter in assuming the measurement noise is Gaussian with the empirically estimated mean and covariance in Section 3.1. To assess the performance, we compare the mean-squared error of the HMB filter with a HMM filter that assumes the underlying process is a Markov chain (instead of a Markov bridge).

Numerical Results: Fig. 4 depicts the results obtained using the above simulation setup. Results show that the proposed HMB filter outperforms the HMM filter for all considered sequence lengths ($T$ values). Thus, the numerical results indicate that the proposed Bayesian filter is capable of accurately estimating the level of segregation in a company-consumer network from noisy sampled edges.

5. CONCLUSION

This paper focuses on a computational sociology problem involving signal processing. We formulated a Markov bridge dynamics based model for how the weights in a social network graph evolve to model segregation. By polling some random pairs of neighbors (i.e. edges) in the network, we formulated an additive Gaussian measurement noise model for the Markov bridge (MB). We proposed a Hidden MB filter to estimate the underlying distance between the communities to provide a real time estimate.
6. REFERENCES

[1] Thomas C Schelling, “Dynamic models of segregation,” Journal of mathematical sociology, vol. 1, no. 2, pp. 143–186, 1971.

[2] Elaine Wallace, Isabel Bui, Leslie de Chernatony, and Michael Hogan, “Who “likes” you... and why? a typology of facebook fans: From “fan”-atics and self-expressives to utilitarians and authentics,” Journal of Advertising Research, vol. 54, no. 1, pp. 92–109, 2014.

[3] Adam Douglas Henry, Pawel Pralat, and Cun-Quan Zhang, “Emergence of segregation in evolving social networks,” Proceedings of the National Academy of Sciences, vol. 108, no. 21, pp. 8605–8610, 2011.

[4] Fabian Baumann, Philipp Lorenz-Spreen, Igor M Sokolov, and Michele Starnini, “Modeling echo chambers and polarization dynamics in social networks,” Physical Review Letters, vol. 124, no. 4, pp. 048301, 2020.

[5] Marloes van Asperen, Pieter de Rooij, and Corné Dijkmans, “Engagement-based loyalty: The effects of social media engagement on customer loyalty in the travel industry,” International Journal of Hospitality & Tourism Administration, vol. 19, no. 1, pp. 78–94, 2018.

[6] Carolyn Heller Baird and Gautam Parasnis, “From social media to social customer relationship management,” Strategy & leadership, 2011.

[7] Carol M Sashi, “Customer engagement, buyer-seller relationships, and social media,” Management decision, 2012.

[8] Miller McPherson, Lynn Smith-Lovin, and James M Cook, “Birds of a feather: Homophily in social networks,” Annual review of sociology, vol. 27, no. 1, pp. 415–444, 2001.

[9] Langford B White and Francesco Caravetta, “Optimal smoothing for finite state hidden reciprocal processes,” IEEE transactions on automatic control, vol. 56, no. 9, pp. 2156–2161, 2011.

[10] M. Fanaswala, V. Krishnamurthy, and L. White, “Destination-aware target tracking via syntactic signal processing,” in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011, pp. 3692–3695.

[11] Fan RK Chung and Fan Chung Graham, Spectral graph theory, Number 92. American Mathematical Soc., 1997.

[12] Buddhika Nettasinghe and Vikram Krishnamurthy, ““what do your friends think?”: Efficient polling methods for networks using friendship paradox,” IEEE Transactions on Knowledge and Data Engineering, 2019.

[13] Buddhika Nettasinghe and Vikram Krishnamurthy, “The friendship paradox: Implications in statistical inference of social networks,” in 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2019, pp. 1–6.

[14] Buddhika Nettasinghe and Vikram Krishnamurthy, “Maximum likelihood estimation of power-law degree distributions using friendship paradox based sampling,” arXiv preprint arXiv:1908.00310, 2019.

[15] Vikram Krishnamurthy, Partially observed Markov decision processes, Cambridge University Press, 2016.

[16] Mustafa Fanaswala, Vikram Krishnamurthy, and Langford White, “Destination-aware target tracking via syntactic signal processing,” in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011, pp. 3692–3695.

[17] Mustafa Fanaswala and Vikram Krishnamurthy, “Detection of anomalous trajectory patterns in target tracking via stochastic context-free grammars and reciprocal process models,” IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 1, pp. 76–90, 2012.

[18] Mustafa Fanaswala and Vikram Krishnamurthy, “Spatiotemporal trajectory models for metalevel target tracking,” IEEE Aerospace and Electronic Systems Magazine, vol. 30, no. 1, pp. 16–31, 2015.

[19] George Stamatescu, Langford B White, and Riley Bruce-Doust, “Track extraction with hidden reciprocal chains,” IEEE Transactions on Automatic Control, vol. 63, no. 4, pp. 1097–1104, 2017.