Learning Common Knowledge Networks Via Exponential Random Graph Models

1st Xueying Liu
Department of Statistics
Virginia Tech
Blacksburg, USA
xliu96@vt.edu

2nd Zhihao Hu
Department of Statistics
Virginia Tech
Blacksburg, USA
huzhihao@vt.edu

3rd Xinwei Deng
Department of Statistics
Virginia Tech
Blacksburg, USA
xdeng@vt.edu

4th Chris J. Kuhlman
Biocomplexity Institute
University of Virginia
Charlottesville, USA
hugo3751@gmail.com

Abstract—Common knowledge (CK) is a phenomenon where each individual within a group knows the same information and everyone knows that everyone knows the information, infinitely recursively. CK spreads information as a contagion through social networks in ways different from other models like susceptible-infectious-recovered (SIR) model. In a model of CK on Facebook, the biclique serves as the characterizing graph substructure for generating CK, as all nodes within a biclique share CK through their walls. To understand the effects of network structure on CK-based contagion, it is necessary to control the numbers and sizes of bicliques in networks. Thus, learning how to generate these CK networks (CKNs) is important. Consequently, we develop an exponential random graph model (ERGM) that constructs networks while controlling for bicliques. Our method offers powerful prediction and inference, reduces computational costs significantly, and has proven its merit in contagion dynamics through numerical experiments.

I. INTRODUCTION

A. Background and Motivation

Common knowledge. Common knowledge (CK) emerges within a group when (i) all individuals possess the same *types* of knowledge $\mathbf{I}$, (ii) each member knows her own information and the information of the other members, and (iii) each member knows that everyone else knows her information. CK is an infinite recursion of knowledge [1], [2]. We study CK among individuals in social networks.

We consider a social network $G(V, E)$ where V is the set of individuals (i.e., nodes) and E is the set of interactions that represent pairwise communication between $i, j \in V$ such that undirected edge $\{i, j\} \in E$. CK enables individuals within a group to coordinate their actions and act simultaneously because they can anticipate each other's actions [3] . We will make this idea concrete in Section II.

There are several reasons for learning *common knowledge networks* (CKNs), i.e., networks that have particular structures that produce CK. First, CK is active in many socio-economic situations, including: (i) forming teams [4], (ii) establishing social norms (e.g., initiating smoking, establishing fear) [5], and (iii) advertising [6]. Second, CK can explain the spread of information through a different mechanism compared to classic mechanisms such as threshold models [7] and susceptible-infectious-recovered (SIR) models [8]. One key difference is that CK enables individuals to coordinate their activation or infection as a group. Another key difference is that CK can *initiate* contagion, in addition to *propagating* it, whereas classic contagion mechanisms only facilitate transmission to a single individual and cannot initiate contagion. Third, CK can explain social phenomena such as preference falsification and fake news spreading [9].

Common knowledge on the Facebook social network. Our work is motivated by a model of CK on the Facebook social media network because of its unique wall or timeline communication mechanism. Figure 1 shows three individuals (1, 2, and 3) connected within a larger social network (see the green edges). Each individual can write to, and read all information on, their own and their distance-1 neighbors' walls (blue boxes); these reading and writing actions are indicated by the black arrows. Therefore, all three individuals, including 1 and 3—who are not directly connected in the network—can directly communicate via the wall of 2 (by writing information to, and reading information from, the wall of 2). In this way, they possess common information $\mathbf{I}=\mathbf{I_1}\cup\mathbf{I_2}\cup\mathbf{I_3}$ via the wall of 2, thus producing CK. The infinite recursion of "knowing that others know" is produced because since 1 sees 3's information $\mathbf{I_3}$ on 2's wall, 1 knows that 3 can see 1's information $\mathbf{I_1}$ on the same wall. Therefore, each of 1 and 3 knows the other's information, and each knows that the other knows that they know. It is this special feature of Facebook walls that enables nodes at distance-2 to directly communicate and form CK.

This leads to a unique CK-based model of contagion [10] which is a coordination game that we formally present in Section II, called CKF (common knowledge on Facebook). Central to this work, it is shown in [10] that the biclique (i.e., complete bipartite graph, where each node in one bipartition is connected to every node in the other bipartition) is the network substructure that produces CK among a collection of
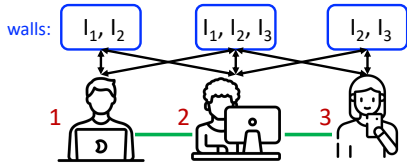
Fig. 1: Illustration of three nodes (node IDs are 1, 2, and 3) in the Facebook social media network arranged in a star graph with green edges, where 2 is the hub node. The Facebook wall of each node or person is above that person (blue box), with the information that is written to, and read from, the wall contained in the box. Information $\mathbf{I_1}$, $\mathbf{I_2}$, and $\mathbf{I_3}$ is the information provided by nodes 1, 2, and 3, respectively. Black arrows indicate who can read from/write to each wall. Persons 1 and 3 can communicate directly through the wall of person 2 even though there is no edge between them.

nodes. (The graph in Figure 1 is a particular type of biclique: a star with hub node 2.) Therefore, it is important to learn to construct CKNs. *The crux of this work is to generate networks with requisite biclique substructures.*

The CK phenomenon has been identified and observed by social scientists (e.g., [3], [6]). CK has also been documented in historical contexts [5]. CK is closely aligned with the concepts of theory of mind [11]; informally, the study of what an individual understands about others' thoughts. Recently, controlled laboratory experiments on human subjects in a Facebook-like [12] and other [13] settings have demonstrated that groups can produce CK.

In this section, we have identified practical situations in which CK is operative and research studies that observe CK. A unique CK-based contagion model has been developed [10] to quantify information contagion, but sensitivity studies of the effect of network structure on contagion spreading have not been done; controlling biclique substructures in networks is the first step in studying these contagions. Hence, there is ample motivation to study the construction of CKNs.

B. Research Questions and Our Contributions

Research questions and technical challenges. Exponential random graph models (ERGMs) are used to construct networks and use relatively simple graph structures for motifs, such as Path$_3$ and triangle substructures [14], [15]. See Figure 2 for an overview of the process. Bicliques, with their large numbers of variants, are comparatively complex structures and have not been used as characterizing subgraphs in ERGMs. (For example, a set of r nodes generates $n_{bic} = \Sigma_{k=1}^{\lfloor r/2 \rfloor} \binom{r}{k}$ possible bicliques; if $r = 5$ then $n_{bic} = 15$.) *Question 1: Can an ERGM be developed to generate CKNs, controlling for bicliques?*

ERGMs do not typically scale well to large graph sizes. Typical undirected network sizes generated with ERGMs are roughly from a few tens of nodes (e.g., [16]–[19]) up to 100,000 nodes in one recent work [20]. *Question 2: Can methods be devised to generate large networks (e.g., 100,000-node) with ERGMs?*

Once CKNs are generated, they must be evaluated to determine whether local network structures from ERGMs are consistent with realistic CKNs. *Question 3: How do ERGM-generated CKNs compare with realistic or mined networks?*

Our contributions. To address the above questions, we develop an ERGM to construct networks while controlling for bicliques. The learning of numbers and distributions of bicliques in social networks can be comprehensively investigated using the proposed method. Moreover, the proposed ERGM method has a parsimonious parametric model formulation with powerful prediction and inference for generating networks. The major contributions of this work are summarized as follows. First, the proposed ERGM for bicliques uses a simple but effective qualitative change statistic, leading to accurate predictions of CKNs. In addition, theoretical properties are established to enable faster calculation of the qualitative change statistics. Second, we use subgraphs of the observed (original) network to construct ERGMs using far less computation time (often, reductions in time by an order of magnitude or more) while preserving comparable prediction accuracy. Third, the effectiveness of the ERGM is evaluated comprehensively in terms of comparisons of degree distributions, k-core distributions, and biclique distributions to demonstrate accurate network predictions. Moreover, we evaluate contagion dynamics on these ERGM-generated networks using an agent-based CKF model to simulate information spreading on Facebook, thereby testing the effect of network structure on dynamics. Contagion spreading on these multiple networks is very similar.

Because no work like ours has been undertaken, as a first step, we focus on generating and evaluating Erdős-Rényi (ER) random graphs. ER graphs represent several real-world systems, such as child friendship networks [21] and communications among people in a work room or classroom [22].

Fig. 2: The ERGM graph generation process where a network instance (at left) is used to estimate an ERGM (center). This model is then used to predict (generate) new graph instances (at right) that are similar to the original graph.

II. PRELIMINARIES

This section presents a formal description of the CKF model [10], focusing on contagion dynamics of bicliques, and an example. A human social network is represented by a communication network $G(V, E)$, where $V = \{1, 2, \ldots, n\}$ is the node set of n people. Each person $i \in V$ is in a state $s_i(t) \in \{0, 1\}$ at each time t. If $s_i(t) = 0$, then person i is in the unactivated state, and if $s_i(t) = 1$, then i is in the activated state. Each node i has a threshold τ_i that indicates its resistance to activation. Given person i's threshold τ_i and the

system state at t, denoted by $s_t = (s_1(t), s_2(t), \ldots, s_n(t))$, her utility is given by

$$U_{it} = \begin{cases} 0 & \text{if } s_i(t) = 0 \\ 1 & \text{if } s_i(t) = 1 \ \wedge \ \left| \{j \backslash i \in V : s_j(t) = 1\} \right| \geq \tau_i \\ -z & \text{if } s_i(t) = 1 \ \wedge \ \left| \{j \backslash i \in V : s_j(t) = 1\} \right| < \tau_i \end{cases},$$
$$(1)$$

where $-z < 0$ is the penalty she gets if she activates and not enough people join her. A person always gets utility 0 by staying in state 0, regardless of what others do since free-riding problems are not considered. When she transitions to the activated state, she gets utility 1 if the total number of *other* people activated at t is at least τ_i. (Note that these "others" do not have to be neighbors of i, as in threshold models, e.g., [7].) We use progressive dynamics [23], such that once in state 1, nodes do not transition back to 0.

Equation (1) clearly shows that node state transitions within a biclique is a coordination game because a collection of individuals is making decisions based on each individual's payoff (utility) and those of others. Also, Equation (1) describes how individuals reason about what they and others will do *in the future* and *simultaneously* make decisions because i, in determining its next state at the *next* time t must infer what other nodes j will also do at the upcoming (next) time t. In contrast, in threshold [24] and SIR [25] models, agents make decisions based on what their neighbors have already done at previous times $t^* < t$.

Simulations of these systems use discrete time to advance the simulation "clock" from time $t = 0$ to t_{max}, where $t \in \mathbb{N}$ can have any unit, although we typically use units of day, i.e., one time tick is one day. At each time, Equation (1) is evaluated for each node $i \in V$ that is in state 0 to determine whether it transitions to state 1; once a node i reaches $s_i(t^*) = 1$, the activated state, the node is assumed to remain active for all $t > t^*$. Two other mechanisms, not associated with bicliques, are also assessed at each time; see [10] for details.

Figure 3 provides an example of CKF model contagion dynamics and a contrast with threshold collective action models. The $K_{3,2}$ biclique, with three nodes in one bipartition (nodes 1 through 3) and two nodes in the second bipartition (nodes 4 and 5), represents five individuals. First, this graph generates CK among all five nodes, as now explained. The star subgraph centered at node 4 generates CK among nodes $\{1, 2, 3, 4\}$ because all of these nodes can read from and write to node 4's wall. Similarly, the star subgraph centered at node 5 generates CK among nodes $\{1, 2, 3, 5\}$ because all of these nodes can read from and write to node 5's wall. To complete CK among all five nodes we need nodes 4 and 5 to know about each other, together with each of the other three nodes: this happens on the walls of nodes 1, 2, or 3. For example, the wall of node 2 generates CK among nodes $\{2, 4, 5\}$. Hence, CK is generated.

The dynamics of contagion that is initiated and spread by CK in Figure 3 is now addressed. The information $\mathbf{I_i}$ that each person (node i) shares on these walls is the triple $(i, \tau_i, s_i(t))$. At $t = 0$, all nodes $i \in V$ are in state $s_i(0) = 0$ (shown as filled red symbols). With all nodes i having $\tau_i = 4$, each i requires four other nodes to activate in order for i to activate, so each node's threshold can be satisfied by the other four members of the CK set. Thus, every i can reason about what it will do next, and because of CK, can reason about what each of the other four nodes will do next. Each node will realize that if all five nodes activate, then everyones' thresholds will be met. Hence, to maximize utility, all nodes will transition to state 1 per Equation (1) in one timestep (green symbols).

This description explains why most contagion simulations use no seed nodes—seed nodes are not needed because CK can initiate contagion where none previously existed. And this is possible because nodes can coordinate their behaviors and act simultaneously.

This behavior can be contrasted with the Granovetter-type threshold model [24], [26]. In that model, a node i in state 0 will transition to state 1 if at least a threshold τ_i number of its neighbors are already in state 1; otherwise i will remain in state 0. Once a node reaches state 1, it remains in state 1 (a progressive model). In our example in Figure 3, no node will transition state $0 \rightarrow 1$ without some seed nodes, which is different from the CKF model. Furthermore, regardless of how many nodes are seeded, there will be no contagion spread because no node has degree of four, so no node has enough neighbors to meet its threshold. Therefore, the only way to have all nodes in state 1 at $t = 1$ is to seed all nodes. Relative threshold models [7], [27] behave in this same way.
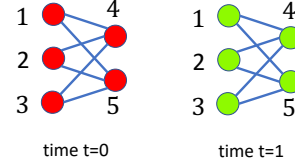

Fig. 3: Example of contagion dynamics on a biclique $K_{3,2}$, using the CKF model. All nodes are assigned thresholds; all five nodes have threshold $\tau = 4$ in this example. For these conditions, all nodes transition to state 1 (in green) in one timestep. In contrast, under the Granovetter threshold model [24], no node will transition state, no matter how many nodes are seeded at time $t = 0$.

The CKF model and the example make clear the prominent role that bicliques play in quantifying CK in Facebook social media networks. Our contributions focus on generating these networks.

III. RELATED WORK

References on common knowledge were given in the Introduction. Here we focus on graph generation methods.

Models for generating property-preserving and property-varying networks enable hypothesis testing, sensitivity analysis, and benchmark testing [15], [28], and exploration of counterfactuals. Several other reasons for researching network generation methods are given in [17]. Studies on the effects of systematically varying network properties include edge shuffling or swapping [29] and using the largest eigenvalue of

an adjacency matrix representation of a graph to characterize epidemic thresholds [30], [31].

There are graph construction methods that use a single graph instance (an observed network) to learn a (stochastic) model. The model can then be used to produce multiple additional instances, presumably from the same family of graphs. Methods include stochastic block models [32] and Kronecker graphs [33]. More recent work includes models that account for node attributes in addition to graph structure [15].

In this work, we focus on a different model of generating multiple graphs from a single graph instance, namely ERGMs [17]. ERGMs have been used to study various problems, including identifying roles of nodes in networks, in addition to their connectivity [34]. Various network substructures (e.g., triangles, paths, stars) have been modeled with ERGMs [14], [16], [17]. None of these structures are as complicated as general bicliques, as we analyze here.

IV. Network Generation Model

In this section, we detail the proposed method of estimating ERGMs that produce biclique structures in social networks.

Consider a set $\mathcal{B}$ of m bicliques $\mathcal{B} = \{B_1, B_2, \ldots, B_m\}$ in graph $G(V, E)$. Each biclique B_j, $1 \leq j \leq m$, consists of two bipartitions $P_{j,1}$ and $P_{j,2}$. Each bipartition is a proper subset of nodes in V of G, with cardinalities $n_{j,1} = |P_{j,1}|$ and $n_{j,2} = |P_{j,2}|$. Let $P_{j,1} = \{k_1, k_2, \ldots, k_{n_{j,1}}\}$ and let $P_{j,2} = \{\ell_1, \ell_2, \ldots, \ell_{n_{j,2}}\}$. The biclique has edges from each node $k_i \in P_{j,1}$, $1 \leq i \leq n_{j,1}$, to every node $\ell_r \in P_{j,2}$, $1 \leq r \leq n_{j,2}$.

Throughout this section, we focus on bicliques where each bipartition has at least two nodes. This is because if one bipartition has one node, then the biclique has a star structure, with the hub node being the sole node in one bipartition and the leaves of the star being the nodes in the other bipartition. These star bicliques are trivial to identify in a graph G because each node $i \in V$ is the hub node of a star subgraph.

A. Exponential Random Graph Model (ERGM)

The ERGM is a probabilistic model for networks. Let us denote a network of n nodes as an adjacency matrix $G^{adj} = (g_{ij}) \in \{0, 1\}^{n \times n}$, where $g_{ij} \in \{0, 1\}$ for all $i, j \in V = \{1, \ldots, n\}$. Here $g_{ij} = 1$ means that there is an undirected edge $e_{ij} = \{i, j\}$ between nodes i and j, while $g_{ij} = 0$ means that the two nodes are not directly connected. We have $g_{ii} = 0$ for all $i \in \{1, \ldots, n\}$. Furthermore, we define the set of all possible simple networks of n nodes as $\mathcal{G}(n) := \{G^{adj} \in \{0, 1\}^{(n \times n)} : g_{ij} \in \{0, 1\}, g_{ii} = 0\}$.

Generally, the probability function of the ERGM is

$$\mathbb{P}_\theta(G^{adj}) = \frac{\exp(\theta^T \Gamma(G^{adj}))}{\sum_{G^* \in \mathcal{G}(n)} \exp(\theta^T \Gamma(G^*))}, \quad (2)$$

where $\theta \in \mathbb{R}^q$ is a q-dimensional vector of parameters. Function $\Gamma(\cdot)$ maps the adjacency matrix G^{adj} to a user-defined q−dimensional vector $\Gamma(G^{adj}) = (\Gamma_1(G^{adj}), \ldots, \Gamma_q(G^{adj}))^T$ of different network statistics. For example, $\Gamma_1(G^{adj})$ can be the number of edges in a network G, and $\Gamma_2(G^{adj})$ can be the number of bicliques in a network. The normalization constant $c(\theta) := \sum_{G^* \in \mathcal{G}(n)} \exp(\theta^T \cdot \Gamma(G^*))$ in Equation (2) ensures that it defines a probability function on $\mathcal{G}$.

Evaluating $c(\theta)$ is very computationally expensive since it needs to enumerate all possible networks of n nodes. Hence, we use the maximum pseudolikelihood estimation (MPLE) [35] for estimating θ, as it is much more approachable. With MPLE, we utilize the pseudolikelihood function $\mathbb{P}_\theta(G^{adj}) = \prod_{ij} \mathbb{P}_\theta(g_{ij}|G^{adj}_{-ij})$ for parameter estimation, where G^{adj}_{-ij} represents the adjacency matrix excluding element g_{ij}. We defined the vector of *change statistics* as $\delta(g_{ij}) = (\delta_1(g_{ij}), \ldots, \delta_q(g_{ij}))^T$ where $\delta_k(g_{ij})$ $(1 \leq k \leq q)$ is the difference in the value of network statistic $\Gamma_k(G)$ when g_{ij} is toggled from 0 to 1 (i.e. $\delta_k(g_{ij}) = \Gamma_k(G^{adj}|g_{ij} = 1) - \Gamma_k(G^{adj}|g_{ij} = 0)$). Note that

$$\mathbb{P}_\theta(g_{ij} = 1|G^{adj}_{-ij}) = \frac{\exp(\theta^T \Gamma(G^{adj}|g_{ij} = 1))}{c(\theta)},$$

and the normalization constant $c(\theta)$ cancels out if we divide the probabilities of g_{ij} being 1 by the probability of being 0:

$$\frac{\mathbb{P}_\theta(g_{ij} = 1|G^{adj}_{-ij})}{\mathbb{P}_\theta(g_{ij} = 0|G^{adj}_{-ij})} = \frac{\exp(\theta^T \Gamma(G^{adj}|g_{ij} = 1))}{\exp(\theta^T \Gamma(G^{adj}|g_{ij} = 0))} = \exp(\theta^T \delta(g_{ij})).$$

Since $g_{ij} \in \{0, 1\}$, $\mathbb{P}_\theta(g_{ij} = 0|G^{adj}_{-ij}) = 1 - \mathbb{P}_\theta(g_{ij} = 1|G^{adj}_{-ij})$, the conditional probability $\mathbb{P}_\theta(g_{ij} = 1|G^{adj}_{-ij})$ can be expressed as a logistic regression of the form

$$\mathbb{P}_\theta(g_{ij} = 1|G^{adj}_{-ij}) = \frac{\exp(\theta^T \delta(g_{ij}))}{1 + \exp(\theta^T \delta(g_{ij}))}.$$

It implies that one can maximize the pseudolikelihood function for the ERGM parameter estimation based on the logistic regression, in which the dependent variable is given by the elements of the adjacency matrix, and the covariates are given by the values of the change statistics corresponding to each element of the adjacency matrix.

B. The Proposed ERGM for Bicliques

An ERGM for bicliques presents several challenges in model formulation, estimation, and computation because bicliques are very complex substructures. Different from substructures like triangles, where there is only one configuration given three nodes, bicliques vary in their sizes and structures. Thus, the identification of node-maximal bicliques (an NP-hard problem) [36] and the calculation of change statistics are computationally challenging. For example, in an $n = 10000$ node graph, there are 10^8 node pairs for evaluating change statistics.

To address the above challenges, we propose a parsimonious ERGM for bicliques. The proposed method has two key strategies. The first is to identify all bicliques in the network. We use an existing code for this [37]. Using the identified bicliques in the network greatly reduces the number of change statistics to be computed. Based on the identified bicliques, our second strategy is to consider a qualitative change statistic for the bicliques in the ERGM. Separating
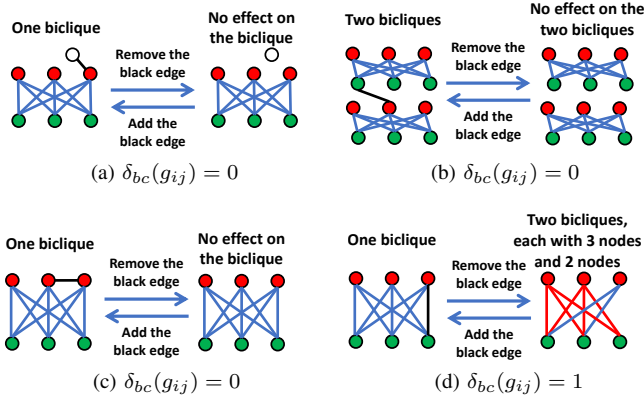
Fig. 4: (a) Removing/adding an edge between two nodes where at least one of them does not belong to any biclique. (b) Removing/adding an edge between two nodes that belong to different bicliques. (c) Removing/adding an edge between two nodes that belong to the same biclique and the same bipartition. (d) Removing/adding an edge between two nodes that belong to the same biclique but different bipartitions. In (d), the node-maximum biclique on the left-hand side has 3 nodes in each bipartition, and there are two node-maximum bicliques with 2 nodes and 3 nodes in the bipartitions on the right-hand side.

these two steps also enables study of multiple change statistics without recomputing the bicliques.

Definition 1. *[**Qualitative change statistic**] For bicliques, let $\delta_{bc}(g_{ij}) \in \{0, 1\}$ be the qualitative change statistic for the ij^{th} dyad g_{ij} in the graph G. $\delta_{bc}(g_{ij}) = 1$ means the existence of an edge between i^{th} and j^{th} nodes does affect the bicliques in G, either in number or size. $\delta_{bc}(g_{ij}) = 0$ means the existence of an edge between i^{th} and j^{th} nodes does not affect the bicliques in the current network.*

First, we illustrate cases to evaluate. Figures 4a through 4d enumerate the cases for adding and removing edges of biclique substructures in a graph. These are the cases that must be addressed in producing an ERGM based on change statistics. Each figure shows one or two bicliques on each side of two arrows. Each biclique on the left-hand side in these examples is a $K_{3,3}$, with a set of three red nodes in one bipartition and a set of three green nodes in the other bipartition. Edges between nodes of the bipartitions are in blue and there is one black edge that is deleted on the right side of the arrows. In the first three of these figures, removal of the black edge produces no change to the number and size of bicliques. Hence, these node pairs have a change statistic of zero, with respect to bicliques. In the last figure, Figure 4d, removal of the black edge causes an increase in the number of bicliques from one to two, and a decrease in their size, from six on the left to two size-five bicliques on the right. Hence, this black edge has a change statistic of one, with respect to bicliques, per Definition 1. Each of these figures also has an arrow from right to left, indicating that in addition to deleting the black

edge, it may also be inserted. The process of deletion and insertion are complements.

From these observations, we formalize the computations of change statistics. First, we define the *overlap* between two bicliques with respect to two nodes i and j.

Definition 2 (**Overlap**). *Consider bicliques B_k and B_ℓ and their two bipartitions $P_{k,1}$ and $P_{k,2}$, and $P_{\ell,1}$ and $P_{\ell,2}$, respectively. Bicliques B_k and B_ℓ overlap with respect to nodes i and j if and only if there exists an assignment of partitions $P_{k,1}$ and $P_{k,2}$, and $P_{\ell,1}$ and $P_{\ell,2}$, such that $P_{\ell,1} \cup \{i\} = P_{k,1}$ and $P_{k,2} \cup \{j\} = P_{\ell,2}$, and there exists a biclique B_{int} under $P_{k,2}$ and $P_{\ell,1}$ as a subgraph of the intersection (or overlap) of B_k and B_ℓ.*

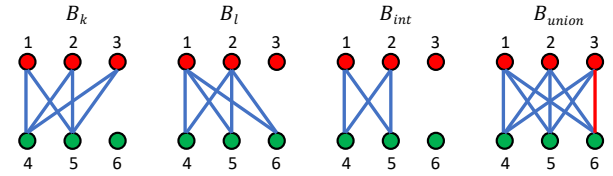A detailed example of the terms in this definition are provided in Figure 5.

Fig. 5: In the definition of overlap, bicliques B_k and B_ℓ overlap with respect to node $i = 3$ and node $j = 6$. Bipartitions for B_k and B_ℓ are $P_{k,1} = \{1, 2, 3\}$ and $P_{k,2} = \{4, 5\}$, and $P_{\ell,1} = \{1, 2\}$ and $P_{\ell,2} = \{4, 5, 6\}$, respectively.

The proposed qualitative change statistic for bicliques avoids the complications of quantifying changes in terms of numbers and sizes of bicliques, but still captures the effect of changes using the binary quantity. It pursues a simple and effective change statistic for bicliques in the ERGM. Moreover, we can establish the following properties for the proposed qualitative change statistic for bicliques, suggested by Figure 4.

Lemma 1. *Given two nodes i, j and an edge $e_{ij} = \{i, j\}$ between them. Removing edge e_{ij} will only affect bicliques that contain e_{ij}. Consider each biclique that contains edge e_{ij}. Removing e_{ij} will affect the number of bicliques and the sizes of the bicliques.*

Lemma 2. *Given two disconnected nodes i, j that belong to two bicliques B_k and B_ℓ, respectively. If there are no bicliques $B_h \in \mathcal{B}$ in which both nodes i and j are a part of (i.e., $\nexists B_h$ such that both $i \in B_h$ and $j \in B_h$), but B_k and B_ℓ overlap, then adding an edge $e_{ij} = \{i, j\}$ between them will create one new biclique from B_k and B_ℓ.*

Now we can establish the following theorem to show the merits of using the proposed qualitative change statistic for bicliques.

Theorem 1 (Qualitative Change Statistic for Bicliques). *Given a graph G with n nodes and m identified bicliques, $\mathcal{B} = \{B_1, B_2, \ldots, B_m\}$. Then the change statistic for dyad g_{ij} can be determined as,*

(a) *If node i and j are in the same biclique but different biparti-tions, then the change statistic for dyad g_{ij} is $\delta_{bc}(g_{ij}) = 1$.*
(b) *If node i and j are in different bicliques B_k and B_ℓ, respec-tively, and B_k and B_ℓ overlap, then $\delta_{bc}(g_{ij}) = 1$.*
(c) *Otherwise, the change statistic $\delta_{bc}(g_{ij}) = 0$.*

Algorithm 1 Qualitative Change Statistics for all Bicliques in a Network

---

Inputs: $G(V, E)$ and set $\mathcal{B}$ of bicliques of G.
Outputs: $\delta_{bc}(g_{ij})$ for all edges $\{i, j\} \in E$.
Steps:
Set $\delta_{bc}(g_{ij}) = 0$ for all edges $e_{ij} \in E$.
for (each $B_k \in \mathcal{B}$) **do**
  for (nodes i and j in different bipartitions of B_k) **do**
    Set the change statistic $\delta_{bc}(g_{ij}) = 1$.
  end for
  for ($B_\ell \in \mathcal{B}$ where B_k and B_ℓ overlap (per Definition 2) with respect to i and j) **do**
    Set the change statistic $\delta_{bc}(g_{ij}) = 1$.
  end for
end for

Algorithm 1 shows the computation of the change statistic for bicliques. The proposed qualitative change statistics for bicliques describe whether the bicliques in a network change given a pair of nodes i, j. Our Algorithm 1 offers an efficient computational method for the qualitative change statistic, surpassing the brute force approach that examines all $\binom{n}{2}$ node pairs for biclique changes.

For the proposed ERGM for bicliques, we fit the model with the network statistics (1) number of edges and (2) number of bicliques in the network. Then the conditional probability is

$$\mathbb{P}_\theta(g_{ij} = 1|G^{adj}_{-ij}) = \frac{\exp[\theta_1 \delta_e(g_{ij}) + \theta_2 \delta_{bc}(g_{ij})]}{1 + \exp[\theta_1 \delta_e(g_{ij}) + \theta_2 \delta_{bc}(g_{ij})]}, \quad (3)$$

where $\delta_e(g_{ij})$ is the change statistic of the edge for the dyad g_{ij}, i.e., the difference in the total number of edges in G when i and j are connected versus disconnected, and $\delta_{bc}(g_{ij})$ is the qualitative change statistic of the biclique. This proposed ERGM is a parsimonious parametric model with only two parameters. Moreover, one can conduct model estimation using a subgraph of the original graph, as shown in the next section. After MPLE, we can make predictions of new networks based on the fitted ERGM in Equation (2) or (3) using Markov chain Monte Carlo (MCMC) techniques [38].

V. NUMERICAL EVALUATION OF ERGMs

We evaluate the proposed ERGM in two ways. First, we quantitatively compare the structural properties of ERGM-generated networks versus those of original graphs. We also present a method to generate these graphs faster using a subgraphing procedure. Second, we compare CKF model contagion dynamics on these networks. The computations determine whether generated networks and original networks produce similar contagion dynamics.

A. Comparison of Graph Structures and Computation

To evaluate the performance of the proposed ERGM model for bicliques, we generate a set of ER graphs G with the same average degree of $d_{ave} = 18$ as shown in Table I.

TABLE I: Erdős-Rényi (ER) random graphs with average degree of $d_{ave} = 18$.

Num Nodes	Num Edges	Edge Prob-ability	Ave Deg.	Max Deg.	Max K-core
2000	18000	9.0×10^{-3}	18	32	12
5000	45000	3.6×10^{-3}	18	39	12
10000	90000	1.8×10^{-3}	18	36	12
20000	180000	9.0×10^{-4}	18	36	12
50000	450000	3.6×10^{-4}	18	39	12
100000	900000	1.8×10^{-4}	18	45	12

For each ER graph G, we first fit the proposed ERGM on it and then make a prediction of a new graph $\hat{G}$. Moreover, we extract one instance of random subgraphs $G_{0.2}, \dots, G_{0.9}$ from G where G_s has a fraction s of the number of nodes of G, determined uniformly at random. Then for each subgraph G_s, $s = 0.2, \dots, 0.9$, we fit the proposed ERGM and predict a new graph $\hat{G}_s$ with the same size (i.e., the same number of nodes) as G. For convenience, we denote $\hat{G}_{1.0}$ for $\hat{G}$. With these graphs, three types of comparisons between original graphs G and ERGM-generated graphs $\hat{G}_s$ are presented next.

We investigate whether the proposed ERGM estimated from subgraphs of G can capture the characteristics of the original graph G. Figure 6 shows the numbers of edges and bicliques in the ERGM-predicted graphs $\hat{G}_s$, $s = 0.2, ..., 1.0$, in comparison with those from the original graph G. For the results in Figure 6 for all network sizes, it is seen that all $\hat{G}_s$ produce a similar number of edges as the original graph G. When $s \geq 0.5$, the $\hat{G}_s$ also give a similar number of bicliques as in the original graph G. For the 20000-, 50000- and 100000-node ER graphs, subgraphs G_s of all fractions produce similar numbers of edges and bicliques as those in the original ER graphs. These results imply that the proposed ERGM can make accurate predictions of graphs. More specifically, the results imply, first, that the proposed ERGM with qualitative change statistics can generate a network $\hat{G}_{1.0}$ with similar numbers of edges and bicliques by using the original graph to fit a model (i.e., use the original G as input to construct an ERGM and then produce $\hat{G} \equiv \hat{G}_{1.0}$). And second, that subgraphs of G can be used to construct ERGMs that then can be used to produce $\hat{G}_s$ that have similar numbers of edges and bicliques as the original G, particularly for $n \geq 20000$ nodes.

We also evaluate our ERGMs using different network met-rics. Specifically, we compare four properties of the related networks: degree distribution, K-core distribution, biclique size distribution, and node participation distribution (i.e., the distribution of number of bicliques that nodes participate in). Figures 7a through 7c and Table II show results for the 50000-node $d_{ave} = 18$ ER graph G in comparison with those from $\hat{G}_{0.2}$, $\hat{G}_{0.5}$, and $\hat{G}_{1.0}$. Figure 7d provides the node
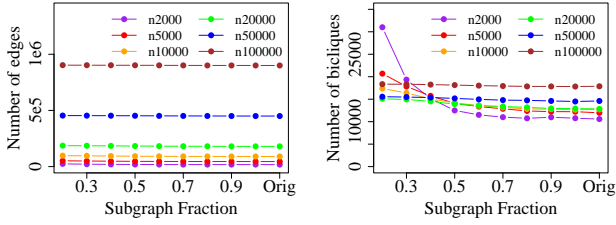
(a) Subgraph Simulation Edges   (b) Subgraph Simulation Bicliques

Fig. 6: (a) Number of edges, and (b) number of bicliques, in ERGM-generated graphs $\hat{G}_s$, $s = 0.2, \ldots, 1.0$, using subgraphs of the original graphs G ("Orig") from Table I.

TABLE II: Kullback–Leibler (K-L) divergence of the node participation distribution and biclique size distribution of each generated graph $\hat{G}_{1.0}$, $\hat{G}_{0.5}$, and $\hat{G}_{0.2}$ compared to those of the original $n = 50000$ graph G. Mean and standard deviation are calculated for 5 replications of graph generation procedure.

Graph	Node Participation Distribution		Biclique Size Distribution	
	Mean	SD	Mean	SD
$\hat{G}_{1.0}$	7.8×10^{-5}	3.3×10^{-5}	4.3×10^{-6}	4.8×10^{-6}
$\hat{G}_{0.5}$	6.8×10^{-4}	9.6×10^{-5}	7.8×10^{-6}	1.1×10^{-5}
$\hat{G}_{0.2}$	1.7×10^{-3}	2.0×10^{-4}	1.7×10^{-6}	2.0×10^{-6}

participation distribution for the 100000-node ER graph. It is seen that $\hat{G}_{1.0}$ can produce similar characteristics as observed in G. Moreover, the data reveal that the proposed ERGM based on subgraphs, i.e., $\hat{G}_{0.2}$, $\hat{G}_{0.5}$, also retains the essential properties and biclique structure of the original graph G. In Table II, the K-L divergence values nearing zero indicate a strong similarity between the distributions of G and $\hat{G}_s$.

These results for $\hat{G}_{0.2}$, $\hat{G}_{0.5}$ are important for another reason, beyond accuracy. For $G_{0.2}$ and $G_{0.5}$, ERGM estimation times (see "Estimation Process" in Figure 2) are only 0.7% and 2% of that for the original 50000-node graph (Figure 8a). The time needed for finding bicliques in subgraphs also decreases significantly, taking only 2% and 17% of the time compared to the original 50000-node graph (Figure 8b). These fractions are even smaller for the 100000-node ER graph. Therefore, our proposed ERGM with subgraphs offers both accurate prediction and computational efficiency.

B. Common Knowledge Simulations

The final part of validation of our ERGM is to compare contagion dynamics that are produced on sets of four networks (the originally constructed network and three ERGM-generated networks). Here we use the CKF contagion model [10] with the focus on the CK mechanism, as in Section II. Our simulation procedures are described in [10].

Agent-based simulations (ABSs) were performed on G, $\hat{G}_{0.2}$, $\hat{G}_{0.5}$, and $\hat{G}_{1.0}$ for both $n = 20000$ and 50000 node graphs. Our focus was on bicliques where each bipartition has a minimum of two nodes, matching the type predicted by the ERGM. We vary node thresholds τ_i and online probabilities

(a) Degree Distribution   (b) K-core distribution

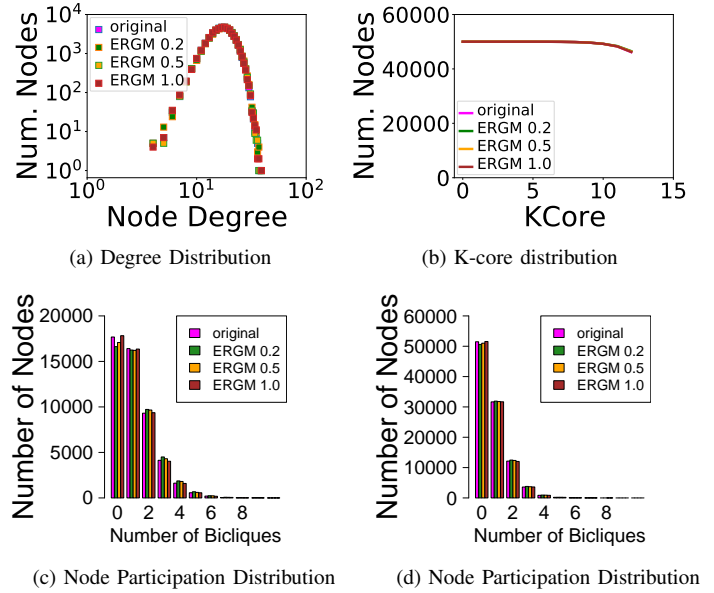(c) Node Participation Distribution   (d) Node Participation Distribution

Fig. 7: Structural analysis results for two networks. For the 50000-node network: (a) degree distributions, (b) K-core distributions, and (c) node participation distributions. (d) Node participation distributions for the $n = 100000$ node network. Each bar depicts the count of nodes that participate in a specific number of bicliques. "Original" is the network provided in Table I, and the three ERGM-generated graphs $\hat{G}_{0.2}$, $\hat{G}_{0.5}$, and $\hat{G}_{1.0}$ are generated using subgraphs of the original graph with fractions of graph nodes of 0.2, 0.5, and 1.0.
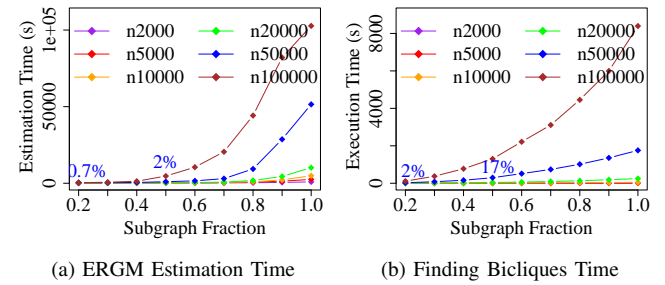

(a) ERGM Estimation Time   (b) Finding Bicliques Time

Fig. 8: (a) ERGM estimation times and (b) times for finding bicliques in average degree $d_{ave} = 18$ ER graphs between 2000 and 100000 nodes, and their subgraphs.

$p_{o,i}$ ($p_{o,i}$ is the probability that an agent is actively participating on Facebook on a particular day of the simulation), conducting 100 simulation instances for each condition to evaluate stochasticity. Each instance progressed from day 0 to day 50, advancing one day at a time.

Figure 9 contains results. The plots are contagion histories where the x-axis is time, in days, and the y-axis is the cumulative fraction of agents that are in state 1 at that time. Various p_o are used, per the legends. In each plot, there are four curves of each color; each curve of one color corresponds to a different graph and is the average of 100 simulation instances. The tight groupings of the four curves of one color indicate that time-wise contagion dynamics are very similar

(a) $n = 20000$, $K_{i,j}$ bicliques
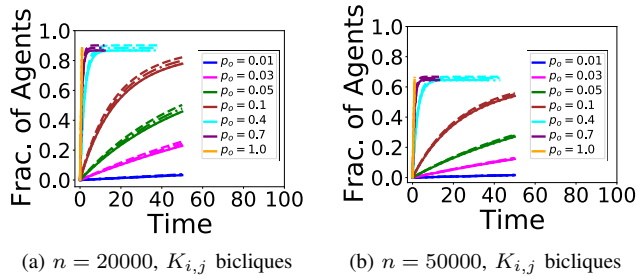
(b) $n = 50000$, $K_{i,j}$ bicliques

Fig. 9: ABS results for the CKF model using only bicliques where each bipartition is of size at least two, denoted $K_{i,j}$. Fraction of agents (nodes) in state 1 is a function of time in days. The data span $t = 0$ to 50, but the x-axes extend to 100 to show legends on the plots without overlapping the curves. (a) and (b) use $n = 20000$ and $n = 50000$ node graphs, respectively, where $\tau_i = 1$ for all nodes and p_o varies per the legend. In each plot, there are four curves for each color, corresponding to ERGM 0.2 (dashed); ERGM 0.5 (dash-dot); ERGM 1.0 (dot); and the original graph (solid).

on the four graphs (the mean absolute differences between the 100 simulation instances of G and each of $\hat{G}_{0.2}$, $\hat{G}_{0.5}$, and $\hat{G}_{1.0}$ are between 0.003 and 0.043). The two plots include only bicliques with at least two nodes per bipartition; the $p_o = 1.0$ curve in Figure 9b shows that about 30% of nodes are not in these bicliques and so cannot change state to 1.

VI. SUMMARY

In this work, we develop exponential random graph models (ERGMs) on the biclique substructures to learn common knowledge networks. This is the first work that uses ERGMs in controlling biclique structures in networks, motivated by the importance of studying the social behavior of common knowledge. Future work includes extending the model and analyses to other classes of networks, and to compare our results with other methods that produce these CKFs.

REFERENCES

[1] R. Aumann, "Agreeing to disagree," *Ann. of Stat.*, vol. 4, pp. 1236–1239, 1976.
[2] R. Lewis, *Convention: A Philosophical Study*. Harvard U. Press, 1969.
[3] M. S.-Y. Chwe, "Structure and strategy in collective action," *Am. J. of Soc.*, vol. 105, pp. 128–156, 1999.
[4] L. A. DeChurch and J. R. Mesmer-Magnus, "The cognitive underpinnings of effective teamwork: A meta-analysis," *Journal of Applied Psychology*, vol. 95, no. 1, pp. 32–53, 2010.
[5] D. Roberts, *The Pueblo Revolt: The Secret Rebellion That Drove the Spaniards Out of the Southwest*. Simon & Schuster, 2004.
[6] M. S.-Y. Chwe, "Culture, circles, and commercials: Publicity, common knowledge, and social coordination," *Rationality and Society*, vol. 10, no. 1, pp. 47–75, 1998.
[7] D. Watts, "A Simple Model of Global Cascades on Random Networks," *PNAS*, vol. 99, no. 9, pp. 5766–5771, 2002.
[8] K. Lum, S. Swarup *et al.*, "The contagious nature of imprisonment: An agent-based model to explain racial disparities in incarceration rates," *Journal of The Royal Society Interface*, vol. 11, no. 98, 2014.
[9] T. Kuran, "The east european revolution of 1989: Is it surprising that we were surprised?" *The American Economic Review*, vol. 81, no. 2, pp. 121–125, 1991.
[10] G. Korkmaz, C. J. Kuhlman, A. Marathe *et al.*, "Collective action through common knowledge using a Facebook model," in *AAMAS*, 2014.
[11] D. Premack and G. Woodruff, "Does the chimpanzee have a theory of mind?" *Behavioral and Brain Sciences*, vol. 1, no. 4, pp. 515–526, 1978.
[12] G. Korkmaz, M. Capra, A. Kraig, K. Lakkaraju, C. J. Kuhlman, and F. Vega-Redondo, "Coordination and common knowledge on communication networks," in *Proceedings of the AAMAS Conference*, 2018, July 10-15, Stockholm, Sweden, 1062–1070.
[13] K. A. Thomas, P. DeScioli, O. S. Haque, and S. Pinker, "The psychology of coordination and common knowledge," *Journal of Personality and Social Psychology*, vol. 107, no. 4, pp. 657–676, 2014.
[14] M. Morris, M. S. Handcock, and D. R. Hunter, "Specification of exponential-family random graph models: Terms and computational aspects," *Journal of Statistical Software*, vol. 24, no. 4, p. 1–24, 2008.
[15] P. Robles, S. Moreno, and J. Neville, "Sampling of attributed networks from hierarchical generative models," in *KDD*, 2016, pp. 1155–1164.
[16] T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, "New specifications for exponential random graph models," *Sociological Methodology*, vol. 36, pp. 99–153, 2006.
[17] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p*) models for social networks," *Social Networks*, vol. 29, pp. 173–191, 2007.
[18] A. Caimo and N. Friel, "Bayesian model selection for exponential random graph models," *Social Networks*, vol. 35, pp. 11–24, 2013.
[19] A. J. Slaughter and L. M. Koehly, "Multilevel models for social networks: Hierarchical Bayesian approaches to exponential random graph modeling," *Social Networks*, vol. 44, pp. 334–345, 2016.
[20] A. Stivala, G. Robins, and A. Lomi, "Exponential random graph model parameter estimation for very large directed networks," *PLoS ONE*, vol. 15, pp. e0 227 804–1–e0 227 804–21, 2020.
[21] K. M. Harris, "The national longitudinal study of adolescent health (add health), waves iii," Carolina Population Center, UNC, Tech. Rep., 2008.
[22] C. L. Barrett, R. J. Beckman, M. Khan *et al.*, "Generation and analysis of large synthetic social contact networks," in *Winter Simulation Conference (WSC)*, 2009, pp. 1003–1014.
[23] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *KDD*, 2003, pp. 137–146.
[24] M. Granovetter, "Threshold Models of Collective Behavior," *American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.
[25] H. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.
[26] T. Schelling, *The Strategy of Conflict*. Harvard U. Press, 1960.
[27] D. Centola and M. Macy, "Complex Contagions and the Weakness of Long Ties," *Am. J. of Soc.*, vol. 113, no. 3, pp. 702–734, 2007.
[28] N. Contractor, S. Wasserman, and K. Faust, "Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example," *Academy of Management Journal*, vol. 31, pp. 681–703, 2006.
[29] S. Eubank, A. Vullikanti, M. Khan, M. Marathe, and C. Barrett, "Beyond degree distributions: Local to global structure of social contact graphs," in *SBP*, 2010, pp. 1–8.
[30] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Transactions on Information and System Security*, vol. 10, no. 4, pp. 13:1–13:26, 2008.
[31] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos, "Threshold conditions for arbitrary cascade models on arbitrary networks," in *ICDM*, 2011, pp. 537–546.
[32] C. Seshadhri, T. G. Kolda, and A. Pinar, "Community structure and scale-free collections of erdős-rényi graphs," *Physical Review E*, vol. 85, no. 5, p. 056109, 2012.
[33] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, "Kronecker graphs: An approach to modeling networks," *J. Mach. Learn. Res.*, vol. 11, p. 985–1042, 2010.
[34] M. Salter-Townshend and T. B. Murphy, "Role analysis in networks using mixtures of exponential random graph models," *Journal of Computational and Graphical Statistics*, vol. 24, no. 2, pp. 520–538, 2015.
[35] D. Strauss and M. Ikeda, "Pseudolikelihood estimation for social networks," *Journal of the American statistical association*, vol. 85, no. 409, pp. 204–212, 1990.
[36] E. L. Lawler, J. K. Lenstra, , and A. H. G. R. Kan, "Generating all maximal independent sets: Np-hardness and polynomial-time algorithms," *SIAM J. on Comp.*, vol. 9, pp. 558–565, 1980.
[37] G. Liu, K. Sim, and J. Li, "Efficient mining of large maximal bicliques," in *LNCS 4081, Conf. DaWak*, 2006, pp. 437–448.
[38] W. R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov chain Monte Carlo in practice*. CRC press, 1995.