Approximate Counting of Graphical Models Via MCMC Revisited

Dag Sonntag [a], Jose M. Peña [a] and Manuel Gómez-Olmedo [b]

[a] ADIT, IDA, Linköping University, Sweden
[b] Dept. Computer Science and Artificial Intelligence, University of Granada, Spain
dag.sonntag@liu.se,jose.m.pena@liu.se,mgomez@decsai.ugr.es

Abstract. We apply MCMC sampling to approximately calculate some quantities, and discuss their implications for learning directed and acyclic graphs (DAGs) from data. Specifically, we calculate the approximate ratio of essential graphs (EGs) to DAGs for up to 31 nodes. Our ratios suggest that the average Markov equivalence class is small. We show that a large majority of the classes seem to have a size that is close to the average size. This suggests that one should not expect more than a moderate gain in efficiency when searching the space of EGs instead of the space of DAGs. We also calculate the approximate ratio of connected EGs to connected DAGs, of connected EGs to EGs, and of connected DAGs to DAGs. These new ratios are interesting because, as we will see, the DAG or EG learnt from some given data is likely to be connected. Furthermore, we prove that the latter ratio is asymptotically 1.
Finally, we calculate the approximate ratio of EGs to largest chain graphs for up to 25 nodes. Our ratios suggest that Lauritzen-Wermuth-Frydenberg chain graphs are considerably more expressive than DAGs. We also report similar approximate ratios and conclusions for multivariate regression chain graphs.

1 Introduction

Graphical models are a formalism to represent sets of independencies, also known as independence models, via missing edges in graphs whose nodes are the random variables of interest.[1] The graphical models are divided into families depending on whether the edges are directed, undirected, and/or bidirected. Undoubtedly, the most popular family is that consisting of directed and acyclic graphs (DAGs), also known as Bayesian networks. A DAG is a graph that contains directed edges and no directed cycle, i.e. no sequence of edges of the form $X \to Y \to \ldots \to X$. As we will see later, it is well-known that different DAGs can represent the same independence model. For instance, the DAGs $X \to Y$ and $X \leftarrow Y$ represent the empty independence model. All the DAGs representing the same independence model are said to form a Markov equivalence class. Probably the most common approach to learning DAGs from data is that of performing a search in the space

[1] All the graphs considered in this paper are labeled graphs.

of either DAGs or Markov equivalence classes. In the latter case, the classes are typically represented as essential graphs (EGs). The EG corresponding to a class is the graph that has the directed edge $A \rightarrow B$ iff $A \rightarrow B$ is in every DAG in the class, and the undirected edge $A - B$ iff $A \rightarrow B$ is in some DAGs in the class and $A \leftarrow B$ is in some others.[3] For instance, the graph $X - Y$ is the EG corresponding to the class formed by the DAGs $X \rightarrow Y$ and $X \leftarrow Y$. Knowing the ratio of EGs to DAGs for a given number of nodes is a valuable piece of information when deciding which space to search. For instance, if the ratio is low, then one may prefer to search the space of EGs rather than the space of DAGs, though the latter is usually considered easier to traverse. Unfortunately, while the number of DAGs can be computed without enumerating them all,[18] the only method for counting EGs that we are aware of is enumeration. Specifically, Gillispie and Perlman enumerated all the EGs for up to 10 nodes by means of a computer program.[7] They showed that the ratio is around 0.27 for 7-10 nodes. They also conjectured a similar ratio for more than 10 nodes by extrapolating the exact ratios for up to 10 nodes.

Enumerating EGs for more than 10 nodes seems challenging: To enumerate all the EGs over 10 nodes, the computer program of Gillispie and Perlman needed 2253 hours in a "mid-1990s-era, midrange minicomputer". We obviously prefer to know the exact ratio of EGs to DAGs for a given number of nodes rather than an approximation to it. However, an approximate ratio may be easier to obtain and serve as well as the exact one to decide which space to search. Therefore, Peña proposed a Markov chain Monte Carlo (MCMC) approach to approximately calculate the ratio while avoiding enumerating EGs.[15] The approach consisted of the following steps. First, the author constructed a Markov chain (MC) whose stationary distribution was uniform over the space of EGs for the given number of nodes. Then, the author sampled that stationary distribution and computed the fraction R of EGs containing only directed edges (EDAGs) in the sample. Finally, the author transformed this fraction into the desired approximate ratio of EGs to DAGs as follows: Since $\frac{\#EGs}{\#DAGs}$ can be expressed as $\frac{\#EDAGs}{\#DAGs}\frac{\#EGs}{\#EDAGs}$,[2] then we can approximate it by $\frac{\#EDAGs}{\#DAGs}\frac{1}{R}$ where $\#DAGs$ and $\#EDAGs$ can be computed as described in the literature.[18;21] The author reported the so-obtained approximate ratio for up to 20 nodes. The approximate ratios agreed well with the exact ones available in the literature and suggested that the exact ratios are not very low (the approximate ratios were in the range [0.26, 0.27] for 7-20 nodes). This suggests that one should not expect more than a moderate gain in efficiency when searching the space of EGs instead of the space of DAGs. Of course, this is a bit of a bold claim since the gain is dictated by the ratio over the EGs visited during the search and not by the ratio over all the EGs in the search space. For instance, the gain is not the same if we visit the empty

[2] We use the symbol # followed by a family of graphs to denote the cardinality of the family.

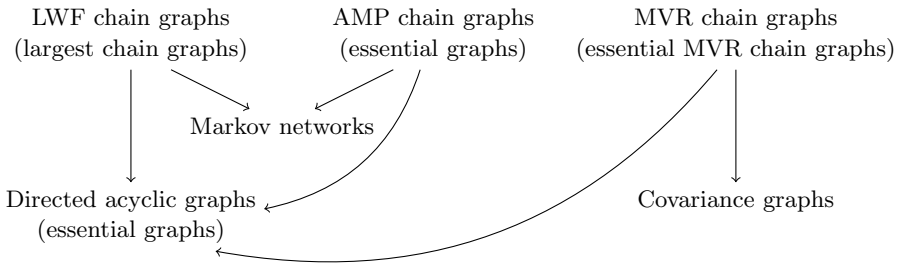

Fig. 1: Families of graphical models considered in this paper. An arrow from a family to another means that every independence model that is representable by the latter family is also representable by the former. For those families where several members may represent the same independence model, we give within parenthesis a unique representative of all such members.

EG, whose ratio is 1, or the complete EG, whose ratio is $1/n!$ for n nodes.[3] Unfortunately, it is impossible to know beforehand which EGs will be visited during the search. Therefore, the best we can do is to draw (bold) conclusions based on the ratio over all the EGs in the search space.

In the work cited above, the author does not only try to elicit which is the most convenient search space for the independence models represented by DAGs, but also to compare the size of this search space with that of a more general family of graphical models known as chain graphs. In other words, the author compares the expressivity of DAGs and chain graphs. Chain graphs (CGs) are graphs with possibly directed and non-directed (i.e. undirected or bidirected) edges, and no semidirected cycle, i.e. no sequence of edges of the form $X \to Y \rightharpoonup \ldots \rightharpoonup X$ where $\rightharpoonup$ is a non-directed edge or a directed edge $\to$ but never $\leftarrow$. Then, CGs extend DAGs and, thus, they can represent at least as many independence models as DAGs. However, unlike DAGs whose interpretation is unique, there are three interpretations of CGs as independence models: The Lauritzen-Wermuth-Frydenberg (LWF) interpretation,[13] the multivariate regression (MVR) interpretation,[5] and the Andersson-Madigan-Perlman (AMP) interpretation.[2] It should be mentioned that no interpretation subsumes any other, i.e. any interpretation can represent independence models that cannot be represented by the other two interpretations.[20] Figure 1 illustrates how the different families of graphical models considered in this paper are related one to another. For any of the three CG interpretations, knowing the ratio of independence models that can be represented by DAGs to independence models that can be represented by

[3] In the latter case, note that there are $n!$ orderings of the nodes in the EG and, thus, there are $n!$ orientations of all the undirected edges in the EG, one per ordering, and none of them has directed cycles.

CGs is a valuable (although not the only) piece of information when deciding which family of graphical models to use. For instance, if the ratio is low, then one may prefer to use CGs rather than DAGs, though the latter are easier to manipulate and reason with. Unfortunately, the only method for computing the fraction that we are aware of is by enumerating all the independence models that can be represented by CGs. This is what Volf and Studený did for LWF CGs by means of a computer program.[22] Specifically, it is well-known that different LWF CGs can represent the same independence model. All such CGs are said to form a Markov equivalence class, which is typically represented by the largest CG (LCG) in the class. The LCG in a class is the CG that has the directed edge $A \to B$ iff $A \to B$ is in every CG in the class.[6] The computer program of Volf and Studený enumerated all the LCGs for up to 5 nodes, which enabled the authors to show the ratio of EGs to LCGs is 1 for 2-3 nodes, 0.93 for 4 nodes, and 0.76 for 5 nodes.

That Volf and Studený ran their computer program only up to 5 nodes indicates that enumerating LCGs is challenging. Therefore, Peña proposed a MCMC approach to approximately calculate the ratio of EGs to LCGs without having to enumerate LCGs.[15] The approach consisted of the following steps. First, the author constructed a MC whose stationary distribution is uniform over the space of LCGs for the given number of nodes. Then, the author sampled this stationary distribution and computed the fraction of the independence models represented by the LCGs in the sample that could also be represented by a DAG. The author reported the so-obtained approximate fraction for up to 13 nodes. The approximate fractions agreed well with the exact ones available in the literature and suggested that the ratio of EGs to LCGs is considerably low (the approximate ratio was 0.04 for 13 nodes). This suggests that one should use CGs instead of DAGs because they are considerably more expressive.

In this paper, we extend the work of Peña in the following four directions.

- We report the approximate ratio of EGs to DAGs for up to 31 nodes. Our ratios are always greater than 0.26, which suggests that the average Markov equivalence class is small.
- We show that a large majority of the Markov equivalence classes of DAGs seem to have a size that is close to the average size.
- We report some new approximate ratios for EGs and DAGs. Specifically, we report the approximate ratio of connected EGs to connected DAGs, of connected EGs to EGs, and of connected DAGs to DAGs. These new ratios are interesting because, as we will see, the DAG or EG learnt from some given data is likely to be connected.
- We report the approximate ratio of EGs to LCGs for up to 25 nodes. We also report similar approximate ratios for MVR CGs. Our results suggest that both LWF CGs and MVR CGs are considerably more expressive than DAGs.

The rest of the paper is organized as follows. Section 2 presents our results for DAGs. Section 3 presents our results for LWF CGs and MVR CGs. Finally,

Table 1: Exact and approximate $\frac{\#EGs}{\#DAGs}$ and $\frac{\#EDAGs}{\#EGs}$.

NODES	EXACT		OLD APPROXIMATE		NEW APPROXIMATE	
	$\frac{\#EGs}{\#DAGs}$	$\frac{\#EDAGs}{\#EGs}$	$\frac{\#EGs}{\#DAGs}$	$\frac{\#EDAGs}{\#EGs}$	$\frac{\#EGs}{\#DAGs}$	$\frac{\#EDAGs}{\#EGs}$
2	0.66667	0.50000	0.66007	0.50500	0.67654	0.49270
3	0.44000	0.36364	0.43704	0.36610	0.44705	0.35790
4	0.34070	0.31892	0.33913	0.32040	0.33671	0.32270
5	0.29992	0.29788	0.30132	0.29650	0.29544	0.30240
6	0.28238	0.28667	0.28118	0.28790	0.28206	0.28700
7	0.27443	0.28068	0.27228	0.28290	0.27777	0.27730
8	0.27068	0.27754	0.26984	0.27840	0.26677	0.28160
9	0.26888	0.27590	0.27124	0.27350	0.27124	0.27350
10	0.26799	0.27507	0.26690	0.27620	0.26412	0.27910
11			0.26179	0.28070	0.26179	0.28070
12			0.26737	0.27440	0.26825	0.27350
13			0.26098	0.28090	0.27405	0.26750
14			0.26560	0.27590	0.27161	0.26980
15			0.27125	0.27010	0.26250	0.27910
16			0.25777	0.28420	0.26943	0.27190
17			0.26667	0.27470	0.26942	0.27190
18			0.25893	0.28290	0.27040	0.27090
19			0.26901	0.27230	0.27130	0.27000
20			0.27120	0.27010	0.26734	0.27400
21					0.26463	0.27680
22					0.27652	0.26490
23					0.26569	0.27570
24					0.27030	0.27100
25					0.26637	0.27500
26					0.26724	0.27410
27					0.26950	0.27180
28					0.27383	0.26750
29					0.27757	0.26390
30					0.28012	0.26150
31					0.27424	0.26710

Section 4 recalls our findings and discusses future work. The paper ends with two appendices devoted to technical details.

2 Directed and Acyclic Graphs

2.1 Average Markov Equivalence Class Size

In this section, we report the approximate average Markov equivalence class size for 2-31 nodes. To be exact, we report the approximate ratio of EGs to DAGs for 2-31 nodes and, thus, the average class size corresponds to the inverse of the ratio. To obtain the ratios, we run the same computer program implementing the MCMC approach described above as Peña does. [15] The experimental settings is also the same for up to 30 nodes, i.e. each approximate ratio reported is based

on a sample of 1e+4 EGs, each obtained as the state of the MC after performing 1e+6 transitions with the empty EG as initial state. For 31 nodes though, each EG sampled is obtained as the state of the MC after performing 2e+6 transitions with the empty EG as initial state. In other words, the ratios reported are based on running in parallel 1e+4 MCs of length 1e+6 for 2-30 nodes and of length 2e+6 for 31 nodes. We elaborate later on why we double the length of the MCs for 31 nodes.

Table 1 presents our new approximate ratios $\frac{\#EGs}{\#DAGs}$ and $\frac{\#EDAGs}{\#EGs}$, together with the old approximate ones and the exact ones available in the literature. The first conclusion that we draw from the table is that the new ratios are very close to the exact ones, as well as to the old ones. This makes us confident on the accuracy of the ratios for 11-31 nodes, where no exact ratios are available in the literature due to the high computational cost involved in calculating them. Another conclusion that we draw from the table is that the ratios seem to be in the range [0.26, 0.28] for 11-31 nodes. This agrees well with the conjectured ratio of 0.27 for more than 10 nodes reported by Gillispie and Perlman.[7] A last conclusion that we draw from the table is that the fraction of EGs that represent a unique DAG, i.e. $\frac{\#EDAGs}{\#EGs}$, is in the range [0.26, 0.28] for 11-31 nodes, a substantial fraction.

Recall from the previous section that we slightly modified the experimental setting for 31 nodes, namely we doubled the length of the MCs. The reason is as follows. We observed an increasing trend in $\frac{\#EGs}{\#DAGs}$ for 25-30 nodes, and interpreted this as an indication that we might be reaching the limits of our experimental setting. Therefore, we decided to double the length of the MCs for 31 nodes in order to see whether this broke the trend. As can be seen in Table 1, it did. This suggests that approximating the ratio for more than 31 nodes will require larger MCs and/or samples than the ones used in this work.

Finally, note that we can approximate the number of EGs for up to 31 nodes as $\frac{\#EGs}{\#DAGs}\#DAGs$, where $\frac{\#EGs}{\#DAGs}$ comes from Table 1 and $\#DAGs$ comes from the literature.[18] Alternatively, we can approximate it as $\frac{\#EGs}{\#EDAGs}\#EDAGs$, where $\frac{\#EGs}{\#EDAGs}$ comes from Table 1 and $\#EDAGs$ can be computed as described in the literature.[21]

2.2 Variability of the Markov Equivalence Class Size

In the previous section, we have shown that $\frac{\#EGs}{\#DAGs}$ is approximately in the range [0.26, 0.28] for 6-31 nodes. This means that the Markov equivalence class size is approximately in the range [3.6, 3.9] on average for 6-31 nodes. In this section, we report on the variability of the class size for 10-32 nodes. Recall that, for n nodes, the class size can vary between 1 and $n!$. However, it has been shown by Gillispie and Perlman by enumerating all the DAGs in all the classes for up to 10 nodes,[7] that a large majority of the classes for a given number of nodes have size ≤ 4.[4] In this section, we provide evidence that that result may hold for

[4] It is difficult to appreciate the exact percentage of classes having size ≤ 4 from their figures, but we estimate that it is not smaller than 70 %.

Table 2: Statistics for the lower and upper bounds of class size.

NODES	STATISTIC	ARROWS	LINES	LOWER BOUND	UPPER BOUND
10	Minimum	8	0	1	1
	Q_1	21	0	1	1
	Q_2	23	1	2	2
	Q_3	25	3	2	6
	Maximum	35	19	120	2073600
15	Minimum	35	0	1	1
	Q_1	51	0	1	1
	Q_2	55	1	2	2
	Q_3	58	3	2	8
	Maximum	74	17	240	1990656
20	Minimum	72	0	1	1
	Q_1	94	0	1	1
	Q_2	99	1	2	2
	Q_3	104	3	2	6
	Maximum	127	22	720	2073600
25	Minimum	117	0	1	1
	Q_1	150	0	1	1
	Q_2	155	1	2	2
	Q_3	161	3	2	6
	Maximum	187	17	120	345600
30	Minimum	90	0	1	1
	Q_1	218	0	1	1
	Q_2	224	1	2	2
	Q_3	231	3	2	6
	Maximum	265	117	39916800	1.996e+50
32	Minimum	58	0	1	1
	Q_1	248	0	1	1
	Q_2	255	1	2	2
	Q_3	263	3	2	8
	Maximum	304	185	8.718e+10	8.827e+83

10-32 nodes too. Therefore, for any number of nodes in the range [2, 32], a large majority of the classes seem to have a size relatively close to the average size which is in the range [3.6, 3.9] and, thus, this average value may be a reasonable estimate of the size of a randomly chosen class.

To arrive at the conclusion above, we ran the same computer program as in the previous section to sample 2e+4 EGs, each obtained as the state of the MC after performing 2e+6 transitions with the empty EG as initial state. The only reason why we doubled parameters as compared to the previous section is because time permitted it. However, time did not permit to compute the sizes of all the classes represented by all the EGs sampled by enumerating all the DAGs in the classes. Therefore, we decided to bound the class sizes instead. A lower bound can be obtained by first finding the largest clique in each connectivity component of the EG and, then, taking the product of the factorials of the sizes of these cliques. An upper bound can be obtained as the product of the factorials of all the cliques in all the connectivity components in the EG. To see

how we arrived at these bounds, note that all the connectivity components of an EG are chordal.[3] Then, we can orient the undirected edges in each connectivity component such that neither immoralities nor directed cycles appear in any former connectivity component.[12] This together with the fact that an EG is a CG that has no induced subgraph of the form $A \to B - C$[3] ensure that neither new immoralities nor directed cycles appear in the resulting directed graph. Then, the resulting directed graph is a DAG that belongs to the class represented by the EG. Specifically, the algorithm arranges all the cliques in a connectivity component in a tree, chooses any of them as the root of the tree, and then orients all the undirected edges in any way that respects the following constraint: $A - B$ cannot get oriented as $A \to B$ if B belongs to a clique that is closer to the root than the clique A belongs to. Therefore, if we choose the largest clique as the root of the clique tree, then we arrive at our lower bound. On the other hand, if we disregard the constraint mentioned when orienting the undirected edges, then we arrive at our upper bound. Our bounds are probably rather loose but, on the other hand, they are easy to compute and, as we will see, tight enough for our purpose. The source code used in this section and the results obtained are publicly available at `https://github.com/mgomez-olmedo/pgm-mcmc`.

From the 2e+4 EGs sampled, we computed the following statistics: Minimum, maximum, and first, second and third quartiles (Q_1, Q_2 and Q_3) of the number of directed edges (arrows), undirected edges (lines), and lower and upper bounds. The results for 10, 15, 20, 25, 30 and 32 nodes can be seen Table 2. The results for the rest of the numbers of nodes considered are very similar to the ones in the table and, thus, we decided to omit them. The first conclusion that we can draw from the table is that whereas Q_1, Q_2 and Q_3 for the number of arrows grow with the number of nodes, Q_1, Q_2 and Q_3 for the number of lines remain low and constant as the number of nodes grows. This implies that, as we can see in the table, Q_1, Q_2 and Q_3 for the lower and upper bounds do not vary substantially with the number of nodes. Recall that all the DAGs in a class only differ in the orientation of some of the lines in the corresponding EG. So, if the EG has few lines, then the class must be small. Note however that the maxima of the lower and upper bounds do vary substantially with the number of nodes, specially between 25 and 30 nodes. We interpret this, again, as an indication that 30 nodes might be the limit of our experimental setting. The second conclusion that we can draw is that Q_1 for the lower and upper bounds is 1. Therefore, $\geq$ 25% of the classes sampled have size 1. This is not a surprise because, as shown in the previous section, $\frac{\#EDAGs}{\#EGs}$ which represents the fraction of classes of size 1 is in the range [0.26, 0.28] for 7-31 nodes. The third conclusion that we can draw is that Q_2 for the lower and upper bounds is 2. Therefore, $\geq$ 50% of the classes sampled have size ≤ 2 and, thus, they are smaller than the average size which is in the range [3.6, 3.9]. The fourth conclusion that we can draw is that Q_3 for the upper bound is 8. Therefore, $\geq$ 75% of the classes sampled have size ≤ 8 and, thus, relatively close to the average size which is in the range [3.6, 3.9]. It is worth mentioning that the last three conclusions agree well with the class size distributions reported by Gillispie and Perlman for up to 10 nodes.[7]

Table 3: Approximate $\frac{\#CEGs}{\#CDAGs}$, $\frac{\#CEGs}{\#EGs}$ and $\frac{\#CDAGs}{\#DAGs}$.

NODES	NEW APPROXIMATE		
	$\frac{\#CEGs}{\#CDAGs}$	$\frac{\#CEGs}{\#EGs}$	$\frac{\#CDAGs}{\#DAGs}$
2	0.51482	0.50730	0.66667
3	0.39334	0.63350	0.72000
4	0.32295	0.78780	0.82136
5	0.29471	0.90040	0.90263
6	0.28033	0.94530	0.95115
7	0.27799	0.97680	0.97605
8	0.26688	0.98860	0.98821
9	0.27164	0.99560	0.99415
10	0.26413	0.99710	0.99708
11	0.26170	0.99820	0.99854
12	0.26829	0.99940	0.99927
13	0.27407	0.99970	0.99964
14	0.27163	0.99990	0.99982
15	0.26253	1.00000	0.99991
16	0.26941	0.99990	0.99995
17	0.26942	1.00000	0.99998
18	0.27041	1.00000	0.99999
19	0.27130	1.00000	0.99999
20	0.26734	1.00000	1.00000
21	0.26463	1.00000	1.00000
22	0.27652	1.00000	1.00000
23	0.26569	1.00000	1.00000
24	0.27030	1.00000	1.00000
25	0.26637	1.00000	1.00000
26	0.26724	1.00000	1.00000
27	0.26950	1.00000	1.00000
28	0.27383	1.00000	1.00000
29	0.27757	1.00000	1.00000
30	0.28012	1.00000	1.00000
31	0.27424	1.00000	1.00000
∞	?	?	≈ 1

In summary, despite the class size can vary between 1 and $n!$ for n nodes, our results suggest that a large majority of the classes have a size relatively close to the average size which is in the range [3.6, 3.9] and, thus, this average value may be a reasonable estimate of the size of a randomly chosen class.

2.3 Average Markov Equivalence Class Size for Connected DAGs

In this section, we report the approximate ratio of connected EGs (CEGs) to connected DAGs (CDAGs). We elaborate below on the relevance of knowing this ratio. For completeness, we also report the approximate ratios of CEGs to EGs, and of CDAGs to DAGs. The approximate ratio of CEGs to CDAGs is computed from the sample obtained in Section 2.1 as follows. First, we compute the ratio R' of EDAGs to CEGs in the sample. Second, we transform this approximate

ratio into the desired approximate ratio of CEGs to CDAGs as follows: Since $\frac{\#CEGs}{\#CDAGs}$ can be expressed as $\frac{\#EDAGs}{\#CDAGs}\frac{\#CEGs}{\#EDAGs}$, then we can approximate it by $\frac{\#EDAGs}{\#CDAGs}\frac{1}{R'}$ where $\#EDAGs$ and $\#CDAGs$ can be computed as described in the literature.[17;21] The approximate ratio of CEGs to EGs is computed directly from the sample. The approximate ratio of CDAGs to DAGs is computed as described in the literature.[17;18]

Gillispie and Perlman state that "the variables chosen for inclusion in a multivariate data set are not chosen at random but rather because they occur in a common real-world context, and hence are likely to be correlated to some degree".[7] This implies that the DAG or EG learnt from some given data is likely to be connected. We agree with this observation, because we believe that humans are good at detecting sets of mutually uncorrelated variables so that the original learning problem can be divided into smaller independent learning problems, each of which results in a CEG. Therefore, although we still cannot say which EGs will be visited during the search, we can say that some of them will most likely be connected and some others disconnected. This raises the question of whether $\frac{\#CEGs}{\#CDAGs} \approx \frac{\#DEGs}{\#DDAGs}$ where DEGs and DDAGs stand for disconnected EGs and disconnected DAGs. Table 3 shows that $\frac{\#CEGs}{\#CDAGs}$ is in the range [0.26, 0.28] for 6-31 nodes and, thus, $\frac{\#CEGs}{\#CDAGs} \approx \frac{\#EGs}{\#DAGs}$. That the two ratios coincide is not by chance because $\frac{\#CEGs}{\#EGs}$ is in the range [0.95, 1] for 6-31 nodes, as can be seen in the table. A problem of this ratio being so close to 1 is that sampling a DEG is so unlikely that we cannot answer the question of whether $\frac{\#CEGs}{\#CDAGs} \approx \frac{\#DEGs}{\#DDAGs}$ with our sampling scheme. Therefore, we have to content with having learnt that $\frac{\#CEGs}{\#CDAGs} \approx \frac{\#EGs}{\#DAGs}$.

From the results in Tables 1 and 3, it seems that the asymptotic values for $\frac{\#EGs}{\#DAGs}$, $\frac{\#EDAGs}{\#EGs}$, $\frac{\#CEGs}{\#CDAGs}$ and $\frac{\#CEGs}{\#EGs}$ should be around 0.27, 0.27, 0.27 and 1, respectively. It would be nice to have a formal proof of these results. In this paper, we have proven a related result, namely that the ratio of CDAGs to DAGs is asymptotically 1. The proof can be found in Appendix A. Note from Table 3 that the asymptotic value is almost achieved for 6-7 nodes already. Our result adds to the list of similar results in the literature, e.g. the ratio of labeled connected graphs to labeled graphs is asymptotically 1.[10]

Note that we can approximate the number of CEGs for up to 31 nodes as $\frac{\#CEGs}{\#EGs}\#EGs$, where $\frac{\#CEGs}{\#EGs}$ comes from Table 3 and $\#EGs$ can be computed as shown in the previous section. Alternatively, we can approximate it as $\frac{\#CEGs}{\#CDAGs}\#CDAGs$, where $\frac{\#CEGs}{\#CDAGs}$ comes from Table 3 and $\#CDAGs$ can be computed as described in the literature.[17]

3 Chain Graphs

Chain graphs (CGs) is a family of graphical models containing two types of edges, directed edges and a secondary type of edge. The secondary type of edge is then used to create components in the graph that are connected by directed edges similarly as the nodes in a DAG. This allows CGs to represents a much

larger set of independence models compared to DAGs, while still keeping some of the simplicity that makes DAGs so useful.

As noted in the introduction, there exist three interpretations of how to read independencies from a CG. However, the three coincide when the CG has only directed edges. Hence, DAGs are a subfamily of the three CG interpretations. It should also be noted that CGs of the LWF interpretation (LWF CGs) and AMP interpretation (AMP CGs) are typically represented with directed and undirected edges, while CGs of the MVR interpretation (MVR CGs) are typically represented with directed and bidirected edges. LWF CGs and AMP CGs containing only undirected edges are also called Markov networks (MNs), whereas MVR CGs containing only bidirected edges are also called covariance graphs (covGs).

In this work, we have chosen to focus on the LWF and MVR interpretations of CGs. More specifically, we study the ratio of independence models that can be represented by MNs and DAGs (resp. covGs and DAGs) to the independence models that can be represented by LWF CGs (resp. MVR CGs). Hereinafter, these ratios are denoted as $R_{LWFtoMNs}$, $R_{LWFtoDAGs}$, $R_{MVRtoCovGs}$ and $R_{MVRtoDAGs}$. Knowing these ratios is a valuable (although not the only) piece of information when deciding which family of graphical models to use. If a ratio is large (close to 1) then the gain of using the more complex CGs is small compared to using the simpler subfamily, while if it is small then one might prefer to use CGs instead of the simpler subfamily.

As mentioned in the introduction, the ratios described above have previously been approximated for LWF CGs for up to 13 nodes.[15] Specifically, the author used a MCMC sampling scheme to sample the space of largest chain graphs (LCGs), because each LCG represents one and only one of the independence models that can be represented by LWF CGs. To our knowledge, we are the first to propose a similar scheme to sample the space of independence models that can be represented by MVR CGs. To do so, we had to overcome some major problems. First of all, there existed no unique representative for each independence model that can be represented by MVR CGs. Hence, one such representative, called essential MVR CGs, had to be defined and characterized. Secondly, no operations for creating a MC over essential MVR CGs existed. This meant that such a set of operations had to be defined so that it could be proven that the stationary distribution of the MC was the uniform distribution. Hence the major contributions in this section, apart from the calculated ratios and their implications, are the essential MVR CGs and the corresponding MC operators.

The rest of the section is structured as follows. First, we define essential MVR CGs and the corresponding MC operators. Then, we move on to the results with a short discussion. Finally, we also discuss some open questions. We have chosen to move theoretical proofs and some additional material for essential MVR CGs to Appendix B.

3.1 Essential MVR CGs

We start by introducing some notation. All graphs in this paper are defined over a finite set of variables V. If a graph G contains an edge between two nodes V_1 and V_2, we denote with $V_1 \to V_2$ a *directed edge*, with $V_1 \leftrightarrow V_2$ a *bidirected edge* and with $V_1 - V_2$ an *undirected edge*. Moreover we say that the edge $V_1 \leftarrow V_2$ has an *arrowhead* towards V_1 and a *non-arrowhead* towards V_2. The *parents* of a set of nodes X of G is the set $pa_G(X) = \{V_1 | V_1 \to V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *spouses* of X is the set $sp_G(X) = \{V_1 | V_1 \leftrightarrow V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *neighbours* of X is the set $nb_G(X) = \{V_1 | V_1 - V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *adjacents* of X is the set $ad_G(X) = \{V_1 | V_1 \to V_2, V_1 \leftarrow V_2$, $V_1 \leftrightarrow V_2$ or $V_1 - V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. A *route* from a node V_1 to a node V_n in G is a sequence of nodes $V_1, \ldots, V_n$ such that $V_i \in ad_G(V_{i+1})$ for all $1 \leq i < n$. A *path* is a route containing only distinct nodes. A path is called a *cycle* if $V_n = V_1$. A path is *descending* if $V_i \in pa_G(V_{i+1}) \cup sp_G(V_{i+1}) \cup nb_G(V_{i+1})$ for all $1 \leq i < n$. A path is *strictly descending* if $V_i \in pa_G(V_{i+1})$ for all $1 \leq i < n$. The *strict descendants* of a set of nodes X of G is the set $sde_G(X) = \{V_n |$ there is a strict descending path from V_1 to V_n in G, $V_1 \in X$ and $V_n \notin X\}$. The *strict ancestors* of X is the set $san_G(X) = \{V_1 | V_n \in sde_G(V_1), V_1 \notin X, V_n \in X\})$. A cycle is called a *semi-directed cycle* if it is descending and $V_i \to V_{i+1}$ is in G for some $1 \leq i < n$. An undirected (resp. bidirected) *component* C of a graph is a maximal (wrt set inclusion) set of nodes such that there exists a path between every pair of nodes in C containing only undirected edges (resp. bidirected edges). A node B is a *collider* in G if G has an induced subgraph of the form $A \to B \leftarrow C$, $A \to B \leftrightarrow C$, or $A \leftrightarrow B \leftarrow C$. Otherwise, B is called *non-collider*. If A and C are adjacent in G, then B is called a *shielded* collider or non-collider. Otherwise, B is called *unshielded* collider or non-collider. A MVR CG is a graph containing only directed and bidirected edges but no semi-directed cycles.

Let X, Y and Z denote three disjoint subsets of V. We say that X is *separated* from Y given Z in a MVR CG G, denoted as $X \perp_G Y | Z$, iff there exists no Z-connecting path between X and Y in G. A path is said to be *Z-connecting* in G iff (1) every non-collider on the path is not in Z, and (2) every collider on the path is in Z or $san_G(Z)$. A node B is said to be a *collider on the path* if the path contains any of the following subpaths: $A \to B \leftarrow C$, $A \to B \leftrightarrow C$, or $A \leftrightarrow B \leftarrow C$. We say that G represents an independence iff the corresponding separation holds in G. The independence model represented by G is the set of independencies whose corresponding separation statements hold in G. As with DAGs and LWF CGs, different MVR CGs can represent the same independence model. All such MVR CGs are said to form a Markov equivalence class. As we have discussed, EGs were introduced to represent Markov equivalence classes of DAGs. EGs were then extended to represent Markov equivalence classes of so-called ancestral graphs.[1] Since ancestral graphs are a superset of MVR CGs, we know on the one hand that all results presented by Ali et al. also hold for MVR CGs, and hence the extended EGs can be used to represent Markov equivalence classes of MVR CGs. On the other hand, we also know that more restrictions

and characteristics might be (and is) possible to assert on the structure of the extended EGs if we only consider MVR CGs rather than ancestral graphs. That is why we define an essential MVR CG as follows.

Definition 1. *A graph G^* is said to be the essential MVR CG of a MVR CG G if it has the same skeleton as G and contains all and only the arrowheads common to every MVR CG in the Markov equivalence class of G.*

From this definition it is clear that an essential MVR CG is an unique representation of a Markov equivalence class of MVR CGs. Note that an essential MVR CG is not actually a MVR CG but might contain undirected edges as well. Each undirected edge implies that there exists two MVR CGs in the Markov equivalence class that have a directed edge between the same two nodes but in opposite directions. Hence, essential MVR CGs contain components containing only undirected edges (undirected components) as well as components containing only bidirected edges (bidirected components). It can be also be shown that no node can be an endnode of both an undirected edge and a bidirected edge and hence that these components are connected to each other with directed edges similarly as the components of any other CG. Using the separation criterion defined above for MVR CGs on essential MVR CGs, we can state the following theorem.

Theorem 1. *An essential MVR CG G^* represents the same independence model as every MVR CG G it is the essential for.*

Now, if we define an indifferent arrowhead as an arrowhead that exists in all the members of a given Markov equivalent class of MVR CGs, then we can give a characterization of essential MVR CGs.

Theorem 2. *A graph G containing bidirected edges, directed edges and/or undirected edges is an essential MVR CG iff (1) it contains no semi-directed cycles, (2) all arrowheads are indifferent, (3) all undirected components are chordal, and (4) all nodes in the same undirected component share the same parents but have no spouses.*

Finally, we can state the following lemmas.

Lemma 1. *There exists a DAG representing the same independence model as an essential MVR CG G^* iff G^* contains no bidirected edges. Moreover, the essential MVR CG is then an EG.*

Lemma 2. *There exists a covG representing the same independence model as an essential MVR CG G^* iff every non-collider in G^* is shielded.*

The proofs of the theorems and lemmas above as well as an algorithm for finding the essential MVR CG for a given MVR CG can be found in Appendix B.

3.2 MC operations

In this section, we propose eight operators that can be used to create a MC whose stationary distribution is the uniform distribution over the space of essential MVR CGs for a given number of nodes. The transition matrix of the MC corresponds to choosing uniformly one of the eight operators and applying it to the current essential MVR CG. Specifically, let G_i be the essential MVR CG before applying operator and G_{i+1} the essential MVR CG after applying the operator. The operators are the following ones.

Definition 2. *MC Operators*

1. *Add Undirected Edge Choose two nodes X and Y in G_i uniformly and with replacement. If the two nodes are non-adjacent in G_i and the graph resulting from adding an undirected edge between them is an essential MVR CG, then let $G_{i+1} = G_i \cup \{X - Y\}$, otherwise let $G_{i+1} = G_i$.*
2. *Remove Undirected Edge Choose two nodes X and Y in G_i uniformly and with replacement. If the two nodes are neighbours in G_i and the graph resulting from removing the undirected edge between them is an essential MVR CG, then let $G_{i+1} = G_i \setminus \{X - Y\}$, otherwise let $G_{i+1} = G_i$.*
3. *Add Directed Edge Choose two nodes X and Y in G_i uniformly and with replacement. If the two nodes are non-adjacent in G_i and the graph resulting from adding a directed edge from X to Y is an essential MVR CG, then let $G_{i+1} = G_i \cup \{X \to Y\}$, otherwise let $G_{i+1} = G_i$.*
4. *Remove Directed Edge Choose two nodes X and Y in G_i uniformly and with replacement. If there exist a directed edge $X \to Y$ between them in G_i and the graph resulting from removing the directed edge between them is an essential MVR CG, then let $G_{i+1} = G_i \setminus \{X \to Y\}$, otherwise let $G_{i+1} = G_i$.*
5. *Add Bidirected Edge Choose two nodes X and Y in G_i uniformly and with replacement. If the two nodes are non-adjacent in G_i and the graph resulting from adding a bidirected edge between them is an essential MVR CG, then let $G_{i+1} = G_i \cup \{X \leftrightarrow Y\}$, otherwise let $G_{i+1} = G_i$.*
6. *Remove Bidirected Edge Choose two nodes X and Y in G_i uniformly and with replacement. If the two nodes are spouses in G_i and the graph resulting from removing the bidirected edge between them is an essential MVR CG, then let $G_{i+1} = G_i \setminus \{X - Y\}$, otherwise let $G_{i+1} = G_i$.*
7. *Add V-collider Chose a node X in G_i uniformly. If $|ad_{G_i}(X)| \geq 2$, let $k = rand(1, |ad_{G_i}(X)|)$ and let V_k be k nodes taken uniformly from $ad_{G_i}(X)$ with replacement. If all edge-endings towards X from every node in V_k are non-arrows, and the graph resulting from replacing these edge-endings with arrows is an essential MVR CG, then let G_{i+1} be such a graph. Otherwise let $G_{i+1} = G_i$.*
8. *Remove V-collider Chose a node X in G_i uniformly. If $|ad_{G_i}(X)| \geq 2$, let $k = rand(1, |ad_{G_i}(X)|)$ and let V_k be k nodes taken uniformly from $ad_{G_i}(X)$ with replacement. If all edge-endings towards X from every node in V_k are arrows, and the graph resulting from replacing these edge-endings with non-arrows is an essential MVR CG, then let G_{i+1} be such a graph. Otherwise let $G_{i+1} = G_i$.*

Table 4: Exact and approximate $R_{LWFtoMNs}$, $R_{LWFtoDAGs}$ and $R_{LWFpureCGs}$.

NODES	EXACT			APPROXIMATE		
	$R_{LWFtoMNs}$	$R_{LWFtoDAGs}$	$R_{LWFpureCGs}$	$R_{LWFtoMNs}$	$R_{LWFtoDAGs}$	$R_{LWFpureCGs}$
2	1.00000	1.00000	0.00000	1.00000	1.00000	0.00000
3	0.72727	1.00000	0.00000	0.71883	1.00000	0.00000
4	0.32000	0.92500	0.06000	0.31217	0.93266	0.05671
5	0.08890	0.76239	0.22007	0.08093	0.76462	0.21956
6				0.01650	0.58293	0.40972
7				0.00321	0.41793	0.57975
8				0.00028	0.28602	0.71375
9				0.00018	0.19236	0.80746
10				0.00001	0.12862	0.87137
11				0.00000	0.08309	0.91691
12				0.00000	0.05544	0.94456
13				0.00000	0.03488	0.96512
14				0.00000	0.02371	0.97629
15				0.00000	0.01518	0.98482
16				0.00000	0.00963	0.99037
17				0.00000	0.00615	0.99385
18				0.00000	0.00382	0.99618
19				0.00000	0.00267	0.99733
20				0.00000	0.00166	0.99834
21				0.00000	0.00105	0.99895
22				0.00000	0.00079	0.99921
23				0.00000	0.00035	0.99965
24				0.00000	0.00031	0.99969
25				0.00000	0.00021	0.99979

Theorem 3. *The MC created from the operators in Definition 2 reaches the uniform distribution over the space of essential MVR CGs for the given number of nodes when the number of transitions goes to infinity.*

3.3 Results

Using the MC operations described by Peña[15] and those described above, LCGs and essential MVR CGs were sampled and the above described ratios calculated. Specifically, 1e+5 LCGs and 1e+5 essential MVR CGs were sampled with 1e+5 transitions between each sample. To check if the independence model represented by a LCG could be represented by a DAG or MN, we made use of the results in the literature.[4] To check if the independence model represented by an essential MVR CG could be represented by a DAG or covG, we made use of Lemma 1 and Lemma 2. The source code used in this section and the results obtained are publicly available at
`www.ida.liu.se/divisions/adit/data/graphs/CGSamplingResources/`.

The calculated approximate ratios can be found in Tables 4 and 5. In addition to these, we also present the exact ratios found through enumeration for up to 5 nodes. Finally we have also added a third ratio, $R_{LWFpureCGs}$ resp.

Table 5: Exact and approximate $R_{MVRtoCovGs}$, $R_{MVRtoDAGs}$, and $R_{MVRpureCGs}$.

NODES	EXACT			APPROXIMATE		
	$R_{MVRtoCovGs}$	$R_{MVRtoDAGs}$	$R_{MVRpureCGs}$	$R_{MVRtoCovGs}$	$R_{MVRtoDAGs}$	$R_{MVRpureCGs}$
2	1.00000	1.00000	0.00000	1.00000	1.00000	0.00000
3	0.72727	1.00000	0.00000	0.72547	1.00000	0.00000
4	0.28571	0.82589	0.10714	0.28550	0.82345	0.10855
5	0.06888	0.59054	0.36762	0.06967	0.59000	0.36787
6				0.01241	0.40985	0.57921
7				0.00187	0.28675	0.71145
8				0.00028	0.19507	0.80465
9				0.00002	0.13068	0.86930
10				0.00000	0.08663	0.91337
11				0.00000	0.05653	0.94347
12				0.00000	0.03771	0.96229
13				0.00000	0.02385	0.97615
14				0.00000	0.01592	0.98408
15				0.00000	0.00983	0.99017
16				0.00000	0.00644	0.99356
17				0.00000	0.00485	0.99515
18				0.00000	0.00267	0.99733
19				0.00000	0.00191	0.99809
20				0.00000	0.00112	0.99888
21				0.00000	0.00073	0.99927
22				0.00000	0.00048	0.99952
23				0.00000	0.00035	0.99965
24				0.00000	0.00017	0.99983
25				0.00000	0.00014	0.99986

$R_{MVRpureCGs}$, describing the ratio of pure LCGs resp. pure essential MVR CGs to all independence models representable by the corresponding interpretation. A pure LCG resp. essential MVR CG represents an independence model that cannot be represented by any DAG or MN resp. DAG or covG. Note that this is not equal to all the independence models that can be represented by LWF CGs minus those that can be represented by DAGs or MNs, since some models can be represented by both DAGs and MNs (and similarly for MVR CGs).

Regarding the accuracy of the approximations, we can see that in both tables the approximations agree well with the exact values. Moreover, plotting the approximated values results in smooth curves, indicating that the LCGs and essential MVR CGs were sampled from an almost uniform distribution. This is further supported by plots of the average number of directed edges, undirected edges or bidirected edges. We omit this results for brevity. For more than 25 nodes, we could however notice inconsistencies in the approximations indicating that not enough MC transitions were performed.

Regarding the approximate ratios themselves, we can see that $R_{LWFtoDAGs}$ and $R_{MVRtoDAGs}$ decrease exponentially when the number of nodes grows. This agrees well with previous results. However, we are the first to identify that this

trend is exponential. Specifically, for more than three nodes, the approximate (and exact) $R_{LWFtoDAGs}$ resp. $R_{MVRtoDAGs}$ almost perfectly follows the curve $9.379 * 0.6512^n$ resp. $5.352 * 0.6614^n$ where n is the number of nodes. Moreover, as $R_{LWFtoMNs}$ and $R_{MVRtoCovGs}$ suggest, MNs and covGs can only represent a very small set of the independence models that CGs and also DAGs can represent. Already for 10 nodes the ratios are $\leq$ 1e-5 and hence, since only 1e+5 graphs were sampled, they are unreliable. Finally, we can see that $R_{LWFpureCGs}$ and $R_{MVRpureCGs}$ grow very fast so that they are ≥ 0.99 for already 15 nodes. Hence, this indicates that there is a large gain in using the more advanced family of CGs compared to DAGs, MNs or covGs, in terms of expressivity.

Note that we can obtain approximate numbers of LCGs and essential MVR CGs for up to 25 nodes by just multiplying the inverse of $R_{LWFtoMNs}$ and $R_{MVRtoCovGs}$ by the corresponding number of independence models that can be represented by MNs and covGs, which is $2^{\frac{n(n-1)}{2}}$ for n nodes. Alternatively, we can multiply the inverse of $R_{LWFtoDAGs}$ and $R_{MVRtoDAGs}$ by the numbers of EGs, which are known exactly for up to 10 nodes or can be estimated for up to 31 nodes as we have described in Section 2.1.

4 Discussion

Gillispie and Perlman have shown that $\frac{\#EGs}{\#DAGs} \approx 0.27$ for 7-10 nodes.[7] We have shown in this paper that $\frac{\#EGs}{\#DAGs}$ is in the range [0.26, 0.28] for 11-31 nodes. These results indicate that the average Markov equivalence class size is in the range [3.6, 3.9] and, thus, one should not expect more than a moderate gain in efficiency when searching the space of EGs instead of the space of DAGs. We have also shown that a large majority of the classes have a size relatively close to the average size which is in the range [3.6, 3.9] and, thus, this average value may be a reasonable estimate of the size of a randomly chosen class. We have also shown that $\frac{\#CEGs}{\#CDAGs}$ is in the range [0.26, 0.28] for 6-31 nodes and, thus, $\frac{\#CEGs}{\#CDAGs} \approx \frac{\#EGs}{\#DAGs}$. Therefore, when searching the space of EGs, the fact that some of the EGs visited will most likely be connected does not seem to imply any additional gain in efficiency beyond that due to searching the space of EGs instead of the space of DAGs.

Some questions that remain open and that we would like to address in the future are checking whether $\frac{\#CEGs}{\#CDAGs} \approx \frac{\#DEGs}{\#DDAGs}$, and computing the asymptotic ratios of EGs to DAGs, EDAGs to EGs, CEGs to CDAGs, and of CEGs to EGs. Recall that in this paper we have proven that the asymptotic ratio of CDAGs to DAG is 1. Another topic for further research would be improving the graphical modifications that determine the MC transitions, because they rather often produce a graph that is not an EG. Specifically, the MC transitions are determined by choosing uniformly one out of seven modifications to perform on the current EG. Actually, one of the modifications leaves the current EG unchanged. Therefore, around 14 % of the modifications cannot change the current EG and, thus, 86 % of the modifications can change the current EG. In our experiments,

however, only 6-8 % of the modifications change the current EG. The rest up to the mentioned 86 % produce a graph that is not an EG and, thus, they leave the current EG unchanged. This problem has been previously pointed out by Perlman.[16] Furthermore, he presents a set of more complex modifications that are claimed to alleviate the problem just described. Unfortunately, no evidence supporting this claim is provided. The problem discussed would not be such if we were to sample the space of DAGs. Specifically, a MCMC approach whose operators only produce DAGs by book-keeping local information has been proposed.[8] As mentioned above, we are not interested in sampling DAGs but independence model represented by DAGs, hence we sample the space of EGs. More recently, He et al. have proposed an alternative set of modifications having a series of desirable features that ensure that applying the modifications to an EG results in a different EG.[11] Although these modifications are more complex than those by Peña[15], the authors show that their MCMC approach is thousands of times faster for 3, 4 and 6 nodes. However, they also mention that it is unfair to compare these two approaches: Whereas 1e+4 MCs of 1e+6 transitions each are run by Peña to obtain a sample, they only run one MC of between 1e+4 and 1e+5 transitions. Therefore, it is not clear how their MCMC approach scales to 10-30 nodes as compared to the one by Peña. The point of developing modifications that are more effective than ours at producing EGs is to make a better use of the running time by minimizing the number of graphs that have to be discarded. However, this improvement in effectiveness has to be weighed against the computational cost of the modifications, so that the MCMC approach still scales to the number of nodes of interest.

In this paper, we have also studied the LWF and MVR interpretations of CGs and shown that only a very small portion of the independence models represented by these can be represented by DAGs, MNs or covGs. More specifically, we have identified that this ratio decreases exponentially when the number of nodes grows. During the process to obtain these results, we have defined and characterized a unique representative for each independence model representable by MVR CGs similar to LCGs for LWF CGs and EGs for DAGs. This allows for future research in what independence models MVR CGs can represent as well how the number of members varies in the different Markov equivalence classes of MVR CGs. In the future, it would also be interesting to look further on how the results for these two interpretations relate to similar results for the AMP interpretation of CGs. Apart from these topics, a future follow-up work could of course consider larger number of nodes. For the MC operations described here for CGs, only about 8 % of the MC transitions were successful, similarly as noted for DAGs. In our experiments, we also observed that the majority of the runtime was spent on checking whether the modified graphs were LCGs or essential MVR CGs. Both these checks can however be done in polynomial time but they can of course be improved, e.g. it may be the case that the check can be done locally depending on the operation applied. One problem with increasing the number of nodes in the experiments is however that $R_{LWFtoDAGs}$ and $R_{MVRtoDAGs}$ decrease exponentially with the number of nodes. Hence, to

get good approximations, the number of graphs sampled would also have to be increased exponentially. For instance, according to the equations fitted to the approximate ratios reported, we can estimate that $R_{MVRtoDAGs}$ is approximately 2e-5 for 30 nodes, and hence only two essential MVR CGs whose independence models could be represented by DAGs would be sampled on average for a sample size of $1e+5$.

Appendix A: Asymptotic Behavior of CDAGs

Theorem 4. *The ratio of CDAGs to DAGs with n nodes tends to 1 as n tends to infinity.*

Proof. Let A_n and a_n denote the numbers of DAGs and CDAGs with n nodes, respectively. Specifically, we prove that $(A_n/n!)/(a_n/n!) \to 1$ as $n \to \infty$. This holds if the following three conditions are met:[24]

(i) $\log((A_n/n!)/(A_{n-1}/(n-1)!)) \to \infty$ as $n \to \infty$,
(ii) $\log((A_{n+1}/(n+1)!)/(A_n/n!)) \geq \log((A_n/n!)/(A_{n-1}/(n-1)!))$ for all large enough n, and
(iii) $\sum_{k=1}^{\infty}(A_k/k!)^2/(A_{2k}/(2k)!)$ converges.

We start by proving that the condition (i) is met. Note that from every DAG G over the nodes $\{v_1, \ldots, v_{n-1}\}$ we can construct 2^{n-1} different DAGs H over $\{v_1, \ldots, v_n\}$ as follows: Copy all the arrows from G to H and make v_n a child in H of each of the 2^{n-1} subsets of $\{v_1, \ldots, v_{n-1}\}$. Therefore,

$$\log((A_n/n!)/(A_{n-1}/(n-1)!)) \geq \log(2^{n-1}/n)$$

which clearly tends to infinity as n tends to infinity.

We continue by proving that the condition (ii) is met. Every DAG over the nodes $V \cup \{w\}$ can be constructed from a DAG G over V by adding the node w to G and making it a child of a subset Pa of V. If a DAG can be so constructed from several DAGs, we simply consider it as constructed from one of them. Let $H_1, \ldots, H_m$ represent all the DAGs so constructed from G. Moreover, let Pa_i denote the subset of V used to construct H_i from G. From each Pa_i, we can now construct $2m$ DAGs over $V \cup \{w, u\}$ as follows: (i) Add the node u to H_i and make it a child of each subset $Pa_j \cup \{w\}$ with $1 \leq j \leq m$, and (ii) add the node u to H_i and make it a parent of each subset $Pa_j \cup \{w\}$ with $1 \leq j \leq m$. Therefore, $A_{n+1}/A_n \geq 2A_n/A_{n-1}$ and thus

$$\log((A_{n+1}/(n+1)!)/(A_n/n!)) = \log(A_{n+1}/A_n) - \log(n+1)$$

$$\geq \log(2A_n/A_{n-1}) - \log(n+1) \geq \log(2A_n/A_{n-1}) - \log(2n) = \log(A_n/A_{n-1}) - \log n$$
$$= \log((A_n/n!)/(A_{n-1}/(n-1)!)).$$

Finally, we prove that the condition (iii) is met. Let G and G' denote two (not necessarily distinct) DAGs with k nodes. Let $V = \{v_1, \ldots, v_k\}$ and $V' =$

$\{v'_1, \ldots, v'_k\}$ denote the nodes in G and G', respectively. Consider the DAG H over $V \cup V'$ that has the union of the arrows in G and G'. Let w and w' denote two nodes in V and V', respectively. Let S be a subset of size $k-1$ of $V \cup V' \setminus \{w, w'\}$. Now, make w a parent in H of all the nodes in $S \cap V'$, and make w' a child in H of all the nodes in $S \cap V$. Note that the resulting H is a DAG with $2k$ nodes. Note that there are k^2 different pairs of nodes w and w'. Note that there are $\binom{2k-2}{k-1}$ different subsets of size $k-1$ of $V \cup V' \setminus \{w, w'\}$. Note that every choice of DAGs G and G', nodes w and w', and subset S gives rise to a different DAG H. Therefore, $A_{2k}/A_k^2 \geq k^2 \binom{2k-2}{k-1}$ and thus

$$\sum_{k=1}^{\infty} (A_k/k!)^2/(A_{2k}/(2k)!) = \sum_{k=1}^{\infty} A_k^2(2k)!/(A_{2k}k!^2)$$

$$\leq \sum_{k=1}^{\infty} ((k-1)!(k-1)!(2k)!)/(k^2(2k-2)!k!^2) = \sum_{k=1}^{\infty} (4k-2)/k^3$$

which clearly converges. □

Appendix B: Proof for Section 3

In this appendix we give the proofs for the theorems defined in Section 3. These proofs do however require some more notation to be defined. Hence this section will start with a notation subsection. This is then followed by the proofs for essential MVR CGs and finally the proof of Theorem 3 is given.

Notation

Note that some of the definitions below were introduced in Section 3.1. We repeat them here for completeness. All graphs in this paper are defined over a finite set of variables V. If a graph G contains an edge between two nodes V_1 and V_2, we denote with $V_1 \to V_2$ a *directed edge*, with $V_1 \leftrightarrow V_2$ a *bidirected edge* and with $V_1 - V_2$ an *undirected edge*. By $V_1 \circ\!\!\to V_2$ we mean that either $V_1 \to V_2$ or $V_1 \leftrightarrow V_2$ is in G. By $V_1 \circ\!\!- V_2$ we mean that either $V_1 \to V_2$ or $V_1 - V_2$ is in G. By $V_1 \circ\!\!-\!\!\circ V_2$ we mean that there exists an edge between V_1 and V_2 in G. A set of nodes is said to be *complete* if there exist edges between all pairs of nodes in the set. Moreover we say that the edge $V_1 \leftarrow V_2$ has an *arrowhead* towards V_1 and a *non-arrowhead* towards V_2.

The *parents* of a set of nodes X of G is the set $pa_G(X) = \{V_1 | V_1 \to V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *children* of X is the set $ch_G(X) = \{V_1 | V_2 \to V_1$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *spouses* of X is the set $sp_G(X) = \{V_1 | V_1 \leftrightarrow V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *neighbours* of X is the set $nb_G(X) = \{V_1 | V_1 - V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$. The *boundary* of X is the set $bd_G(X) = pa_G(X) \cup nb_G(X) \cup sp_G(X)$. The *adjacents* of X is the set $ad_G(X) = \{V_1 | V_1 \to V_2, V_1 \leftarrow V_2, V_1 \leftrightarrow V_2$ or $V_1 - V_2$ is in G, $V_1 \notin X$ and $V_2 \in X\}$.

A *route* from a node V_1 to a node V_n in G is a sequence of nodes $V_1, \ldots, V_n$ such that $V_i \in ad_G(V_{i+1})$ for all $1 \leq i < n$. A *path* is a route containing only distinct nodes. The length of a path is the number of edges in the path. A path is called a *cycle* if $V_n = V_1$. A path is *descending* if $V_i \in pa_G(V_{i+1}) \cup sp_G(V_{i+1}) \cup nb_G(V_{i+1})$ for all $1 \leq i < n$. A path $\pi = V_1, \ldots, V_n$ is *minimal* if there exists no other path π_2 between V_1 and V_n such that $\pi_2 \subset \pi$ holds. The *descendants* of a set of nodes X of G is the set $de_G(X) = \{V_n|$ there is a descending path from V_1 to V_n in G, $V_1 \in X$ and $V_n \notin X\}$. A path is *strictly descending* if $V_i \in pa_G(V_{i+1})$ for all $1 \leq i < n$. The *strict descendants* of a set of nodes X of G is the set $sde_G(X) = \{V_n|$ there is a strict descending path from V_1 to V_n in G, $V_1 \in X$ and $V_n \notin X\}$. The *ancestors* (resp. *strict ancestors*) of X is the set $an_G(X) = \{V_1 | V_n \in de_G(V_1), V_1 \notin X, V_n \in X\}$ (resp. $san_G(X) = \{V_1 | V_n \in sde_G(V_1), V_1 \notin X, V_n \in X\}$). A cycle is called a *semi-directed cycle* if it is descending and $V_i \to V_{i+1}$ is in G for some $1 \leq i < n$.

An undirected (resp. bidirected) *component* C of a graph is a maximal (wrt set inclusion) set of nodes such that there exists a path between every pair of nodes in C containing only undirected edges (resp. bidirected edges). If the type (undirected resp. bidirected) is not specified we mean either type of component. We denote the set of all connectivity components in a graph G by $cc(G)$ and the component to which a set of nodes X belong in G by $co_G(X)$. A *subgraph* of G is a subset of nodes and edges in G. A subgraph of G induced by a set of its nodes X is the graph over X that has all and only the edges in G whose both ends are in X.

In this appendix we deal with three families of graphs; directed acyclic graphs (DAGs), multivariate regression chain graphs (MVR CGs) and joined chain graphs (JCGs). A DAG contains only directed edges and no semi-directed cycles. A MVR CG is a graph containing only directed and bidirected edges but no semi-directed cycles. Joined graphs are graphs created by joining ancestral graphs. [1] JCGs have a similar definition with the exception that the graphs joined must be MVR CGs of the same Markov equivalence class. Hence we can define JCGs as:

Definition 3. *Joined Chain Graph If G_1 and G_2 are two MVR CGs belonging to the same Markov equivalence class then define the joined chain graph $G = G_1 \vee G_2$ to be the graph with the same adjacencies such that, on an edge between X and Y, there is an arrowhead towards X in G if and only if there is an arrowhead towards X in both G_1 and G_2.*

Let X, Y and Z denote three disjoint subsets of V. We say that X is *separated* (in the families of graphs described above) from Y given Z denoted as $X \perp_G Y | Z$ iff there exists no Z-connecting path between X and Y. A path is said to be *Z-connecting* iff (1) every non-collider on the path is not in Z, (2) every collider on the path is in Z or $san_G(Z)$, and (3) no arrowheads meet any undirected edges. A node B is said to be a *collider* in a JCG, MVR CG or DAG G between two nodes A and C on a path if one of following configurations exists in G: $A \circ\!\!\to B \leftarrow\!\!\circ C$ while for any other configuration ($A \circ\!\!- B -\!\!\circ C$,$A \circ\!\!\to B -\!\!\circ C$ or $A \circ\!\!- B \leftarrow\!\!\circ C$) is considered a non-collider. Note that the definition simplifies somewhat for

example MVR CGs since they cannot contain any undirected edges. We also say that a collider resp. non-collider is shielded if A and C is adjacent, otherwise we say that it is unshielded.

The *independence model* M induced by a graph G, denoted as $I(G)$ or $I_{PGM-family}(G)$, is the set of separation statements $X \perp_G Y|Z$ that hold in G according to the interpretation to which G belongs or the subscripted PGM-family. We say that two graphs G and H are *Markov equivalent* (under the same interpretation) or that they are in the same *Markov equivalence class* iff $I(G) = I(H)$. If an arrowhead occurs on an edge between the same nodes in all graphs of a Markov equivalence class we say that it is indifferent, otherwise we say that it is not indifferent. Moreover we also say that an edge is indifferent if it exists in every graph of a Markov equivalence class. Finally we do also define the skeleton of a graph G as a graph with the same structure as G with the exception that all edges have been replaced by undirected edges.

Essential MVR CGs

As noted in section 3.1 an essential MVR CG G^* of a MVR CG G if it has the same skeleton as G and contain all and only the arrowheads that are shared by all MVR CGs in the Markov equivalence class of G. Hence another definition can be that the essential MVR CG G^* of a MVR CG G is the JCG created when all MVR CGs in the Markov equivalence class of G are joined.

We can now go on to prove Theorem 1:

Proof. We know that any other MVR CG G' in the same Markov equivalence class as G must contain the same adjacencies as well as the same unshielded colliders.[23] Since a collider over a node Y between two nodes X and Z is a subgraph of the form $X \circ\!\!\rightarrow Y \leftarrow\!\!\circ Z$ with arrowheads towards Y we know that these arrowheads must be in every MVR CG G' in the same Markov equivalence class as G. Hence we know that the collider also must be in G^*. Since the definition of d-separation is the same for the JCGs and MVR CGs, and no new colliders can be created or removed when graphs of the same Markov equivalence class are joined, we know that the theorem must hold. □

For the proofs for the remainder of the theorems in subsection 3.1 we do however need to show how an essential MVR CG G^* can be found from a MVR CG G. In their work of presenting essential graphs for AGs Ali et al. defines an algorithm for transforming an ancestral graph into its essential graph. Since MVR CGs are a subfamily of AGs this algorithm do of course also work for creating the essential MVR CG of some MVR CG. However, just like for the characteristics some simplifications (and optimizations) can be made if one only considers the more restricted MVR CGs. The goal of the algorithm is to find all indifferent arrowheads in a graph, i.e. all the arrowheads that must exist in all graphs of that Markov equivalent class. This is exactly what is done in (parts of) the PC-algorithm and hence the MVR CG PC-algorithm[19] can be used to transform a MVR CG into its essential graph. If one studies the algorithm given by Ali et al. it can in fact also be seen that if it is simplified to only work for MVR

CGs it becomes similar to the MVR CG PC-algorithm. We have here added the MVR-PC-algorithm (row 2 to 9) slightly modified to find the essential graph:

Transformation Algorithm
Given a MVR CG G the algorithm learns a joined graph G^* that is the essential graph of G

1 Let G^* be the skeleton of G
2 Orient any induced subgraph $X \circ\!\!-\!\!\circ Y -\!\!\circ Z$ in G^* to $X \circ\!\!\rightarrow Y \leftarrow\!\!\circ Z$ iff X and Z form an unshielded collider over Y in G.
3 Apply the rules in Figure 2 onto G^* until no further arrowheads are added.

Algorithm 1: Transformation algorithm

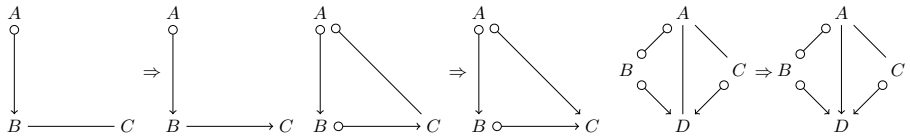

Fig. 2: The rules

Theorem 5. *The transformation algorithm given in Algorithm 1 is correct and always learns the essential MVR CG G^* given an MVR CG G as input.*

Proof. To see that the transformation algorithm is correct we have to show two things. First, that the algorithm given in Algorithm 1 gives the same result as the first 9 lines of the PC-algorithm for MVR CGs[19] when we have a MVR CG G instead of a probability distribution p faithful to G as input. Here we know that line 1 to 7 in the MVR CG PC-algorithm finds the skeleton of G, and hence, since we already have the skeleton of G, can be replaced by line 1 in the transformation algorithm shown here. We can also replace rule 0 in line 8 in the MVR CG PC-algorithm by line 2 in the transformation algorithm since we know that any node B for which rule 0 can be applicable must be an unshielded collider in G. Finally line 9 in the MVR CG PC-algorithm can be replaced by line 3 here since we know that all unshielded colliders are found and orientated in G^* in line 2 in the transformation algorithm. Hence for any triplet of nodes A, B, C, such that the induced subgraph $A \circ\!\!-\!\!\circ B -\!\!\circ$ exists in G^*, B must be in the separator of A and C if it reaches line 3. Hence this prerequisite can be removed from the rules, which are otherwise the same as for the MVR CG PC-algorithm.

Secondly, we must show that first 9 lines of the MVR-PC algorithms gives the essential graph. This follows directly from that all the rules are sound[19] and that any node can be chosen to be node of order 0 when orientating the

remaining undirected edges in line 10 to 14. That the rules are sound means that the arrowheads introduced must exist in every MVR CG of the Markov equivalence class of G. Hence all arrowheads in G^* after line 3 in Algorithm 1 are indifferent in the Markov equivalence class of G. That any node can be chosen to have order 0 means that any remaining undirected edge can be orientated as a directed edge in either direction and the resulting MVR CG can still belong to the same Markov equivalence class if the rest of the undirected edges are oriented appropriately. Hence all undirected edges $X-Y$ have at least one MVR CG G_1 such that $I(G_1) = I(G)$ where the edge $X \to Y$ exists in G_1 and one MVR CG G_2 such that $I(G_2) = I(G)$ where the edge $X \leftarrow Y$ exists in G_2. □

Having this transformation algorithm we can now define some characteristics of essential MVR CGs:

Lemma 3. *For any essential MVR CG G^* we know that:*

1. *All bidirected edges are indifferent in the Markov equivalence class of MVR CGs that G^* is the essential MVR CG for.*
2. *No undirected edge can share an endnode with a bidirected edge.*
3. *An induced subgraph of the form $A \to B-C$ cannot exist in an essential MVR CG. Hence all nodes in any undirected component must share the same parents.*
4. *Any undirected component is chordal.*
5. *No semi-directed cycles can occur.*

Proof. Point 1 follows directly from that all arrowheads are indifferent.

Point 2 follows from existing results.[19]

Point 3 must hold or the first rule in Figure 2 would be applicable which is a contradiction.

Point 4 follows from existing results.[19].

Point 5 must hold since we know that there exist a MVR CG G with the same directed and bidirected edges as G^* but where every undirected edge is made directed such that $I(G) = I(G^*)$. Hence we know that no semi-directed cycle can occur in G^* with only directed and bidirected edges since such a semi-directed cycle then would occur in G which is a contradiction. On the other hand, if an undirected edge $X-Y$ is part of the semi-directed cycle we know that $pa_{G^*}(X) = pa_{G^*}(Y)$ by point 3 above, and hence there must also exist a semi-directed cycle without the undirected edge in G^* because X and Y have no spouses by point 2 above. This reasoning can be repeated for every undirected edge in the semi-directed cycle. This means that if there existed a semi-directed cycle containing undirected edges there also must exist a semi-directed cycle without any undirected edges, which we know cannot be the case. Hence we have a contradiction. □

Which finally allows us to prove Theorem 2:

Proof. We will first prove that if any of the conditions are not fulfilled than the graph cannot be an essential MVR CG. Point 1 follows directly from point 5 in Lemma 3. Point 2 follows from the definition of indifferent arrowheads. Point 3 follows from point 4 in Lemma 3 and point 4 follows from point 2 and 3 in Lemma 3.

To show that the graph must be an essential MVR CG it is enough to show that if the conditions are fulfilled then an MVR CG exists in the same Markov equivalence class since we know that all arrowheads in G are indifferent. Hence we need to show that the undirected edges in G can be oriented as directed edges such that no new unshielded colliders are added or semi-directed cycles are created. Since the undirected components are chordal we know that there exists a way to orient every edge as a directed edge such that no semi-directed cycles or unshielded colliders are added including only the oriented edges. [12] However, we still need to show that no new unshielded colliders or semi-directed cycles are created including the already directed edges in G. To see this note that no new unshielded colliders can be created since $pa_G(X) = pa_G(Y)$ and $sp_G(X) = sp_G(Y) = \emptyset$ for any two X and Y in the same undirected component. In addition, since all nodes in the same undirected component share the same parents it follows that if a semi-directed cycle, including a directed edge in G, would be created when the undirected edges are oriented, then there also must exist a semi-directed cycle in G. From point 1 we do however know that this is not the case, and hence there must exist an MVR CG in the same Markov equivalence class as G for which the undirected edges have been oriented. □

We can also prove Lemma 1 given in section 3:

Proof. Let $\mathcal{G}$ be the Markov equivalence class of G^*. That a DAG $G \in \mathcal{G}$ exists must hold since we know that there exist a MVR CG $G' \in \mathcal{G}$ with the same structure as G^* with the exception that all undirected edges have been replaced by directed edges. Hence G' only contains directed edges and hence is a DAG. That no DAG exists if G^* contains any bidirected edge we simply have to note that DAGs is a subfamily of essential MVR CGs and that all arrowheads are indifferent in an essential MVR CG. Hence there can exist no member in the Markov equivalence class not containing the bidirected edge.

To see that G^* is the essential graph we can note the following. For any MVR CG $G' \in \mathcal{G}$ there must exist a DAG G such that all arrowheads in G exist in G' since G' only can contain bidirected edges where G contains directed edges. Hence it is enough to join every DAG in $\mathcal{G}$ to get the essential MVR CG G^* and hence, since this is a way the essential graph can be created for DAGs, [3] G^* must be the essential graph of $\mathcal{G}$. □

As well as Lemma 2:

Proof. Let $\mathcal{G}$ be the Markov equivalence class of G^*. We know that a covG $G \in \mathcal{G}$ must exist if there exists no non-shielded non-colliders in G^* since for all triplets X, Y and Z, such that G^* the induced subgraph $X \circ\!\!-\!\!\circ Y \circ\!\!-\!\!\circ Z$, X and Z must form an unshielded collider over Z. Hence all edges in G^* can be replaced by bidirected edges to get G such that $I(G) = I(G^*)$. To see that no covG $G \in \mathcal{G}$

can exist if there exist a non-shielded non-collider in G^* let this non-collider be over Y such that the induced subgraph $X \circ\!\!-\!\!\circ Y -\!\!\circ Z$ exists in G^* such that X is a non-descendant of Z. Clearly G must have the same adjacencies as G^* since covGs is a subfamily of essential MVR CGs and hence the induced subgraph $X \circ\!\!-\!\!\circ Y \circ\!\!-\!\!\circ Z$ must exist in G. However, since G only contains bidirected edges this would mean that $X \not\perp_G Z|Y \cup pa_{G^*}(Z)$ while $X \perp_{G^*} Z|Y \cup pa_{G^*}(Z)$ must hold. Hence we have a contradiction. □

Finally we also state one more lemma about the structure of JCGs that is used later in the appendix.

Lemma 4. *The arrowhead on any directed edge that shares an endnode with an indifferent bidirected edge in a JCG must be indifferent itself.*

Proof. This follows from the fact that a bidirected edge cannot share an endnode with an undirected edge in an essential MVR CG as denoted in point 2 in Lemma 3. This means there can exist no directed edge $X \to Y$ in a MVR CG G such that there exist another MVR CG G' of the same Markov equivalence class containing $X \leftarrow Y$ if X or Y is an endnode of a bidirected edge. Hence the arrowhead must be indifferent. □

MCMC operators

The rest of the appendix is devoted to proving Theorem 3. This is performed by first giving the proof of the theorem which then uses lemmas defined later in the appendix. This structure does hopefully allow the reader to get a better understanding why the lemmas are needed and what they will be used for. Hence we first give the proof of Theorem 3:

Proof. Since we know that the possible states are the essential MVR CGs we need to prove two things; first that the Markov chain has a unique stationary distribution and secondly that this is the uniform distribution. The former is proven if we show that the operators have the properties aperiodicity, i.e. that the Markov chain does not end up in the same state periodicly, and irreducibility, i.e. that any state can be reached from any other state.[9] The latter is proven if we can show that the operators have the symmetry property, i.e. that the probability of transition from one state to another is equal to the probability of going to the latter state to the former.[9]

To prove aperiodicy we only need to show that there exists a positive probability of ending up in the same state in two consequent transitions (given that irreducibility holds).[9] This obviously holds since there must always exist an operation that for a set of certain nodes will result in that $G_{i+1} = G_i$ (i.e. if "Add undirected edge" operation for two nodes A and B is possible, then the "Remove undirected edge" operation for the two nodes A and B cannot be possible).

To prove irreducibility we need to show that we with these operators can reach any essential MVR CG G' from any other essential MVR CG G''. It is here enough to show that we from the empty graph $G_\emptyset$ can reach any essential

MVR CG, since we know that any operation is reversible, i.e. it can be undone by some other operation. To show this we can base our reasoning on the following: let G^* be the essential MVR CG we want to reach through a set of operations. It must then exist another MVR CG G_u^* that has the same structure as G^* with the exception that G_u^* contains no undirected edge that exists in G^*. Lemma 5 states that we can remove edge by edge to reach G_u^* from G^* by only removing undirected edges one by one, and due to reversibility we then also know G^* is reachable by adding undirected edges one by one to G_u^*. Hence we now only need to show that G_u^* is reachable from the empty graph $G_\emptyset$. Lemma 7 then states that an essential MVR CG $G_{C_{i+1}}$ is reachable, with the operations described in Definition 2, from another essential MVR CG G_{C_i} if $G_{C_{i+1}}$ and G_{C_i} have the same structure with the exception that $G_{C_{i+1}}$ contains one more bidirected component C such that $ch_{G_{C_{i+1}}}(X) = \emptyset$. Hence we know that we can add components one by one such that we achieve a set of essential MVR CGs $G_{C_0}, G_{C_1}, ..G_{C_n}$ where $G_{C_0} = G_\emptyset$ and $G_{C_n} = G_u^*$ and that all the necessary transformations is possible if the operations in Definition 2 is chosen in the right order. Hence we have irreducibility.

Finally we have to prove symmetry. This must also hold since the probability of choosing an operation o for an essential MVR CG G_i that transforms G_i to an essential MVR CG G_{i+1} must be equal to choosing an operation o' for G_{i+1} such that o' transforms G_{i+1} to G_i. To see this note that the probability of choosing any of the first six operators for a certain set of nodes is $\frac{1}{8} * \frac{1}{n^2}$ for any essential MVR CG. Hence the "remove" operator for a certain kind of edge must have an equal probability of being chosen in G_{i+1} as the "add" operator for the same kind of edge in G_i for these operators. For the seventh and eight operation the probability is harder to state. However, let G_i be the essential MVR CG before a "add V-collider operation" over a node X, where k non-arrow edge-endings have been changed to arrows, and G_{i+1} the essential MVR CG after the operation. Then the probability of the "add V-collider operation" in G_i is $\frac{1}{8} * \frac{1}{n} * \frac{1}{|ad_{G_i}(X)|^k}$. If we then study the reverse operator "Remove V-collider" in G_{i+1} we can see that the probability of transforming G_{i+1} to G_i is $\frac{1}{8} * \frac{1}{n} * \frac{1}{|ad_{G_{i+1}}(X)|^k}$. Since we know that $ad_{G_i}(X) = ad_{G_{i+1}}(X)$ we can deduce that the probability must be equal in both cases. Hence we have symmetry for the set of operations. □

Lemma 5. *In a chordal undirected graph G, with at least one edge, it is always possible to remove one edge so that the resulting graph G' is still chordal.*

Proof. From Jensen and Nielsen[14] we know that there exists at least two simplicial nodes in G such that, since G contains an edge, have at least one neighbour. A simplicial node is a node such that the set of neighbours of the node is complete. Let X be a simplicial node that has a neighbour Y. Now assume that removing $X-Y$ from G creates a nonchordal cycle. Then that cycle must be of the form $V_1, \ldots, V_n = V_1$ with $V_2 = X$ such that V_1 and V_3 is not adjacent (or the cycle must also be non-chordal in G which cannot be the case). However, that V_1 and V_3 is not adjacent is a contradiction since X is simplicial, and hence $G \setminus \{X-Y\}$ must be chordal. □

Lemma 6. *In an essential MVR CG G^* with a bidirected component C there exists an unshielded collider, between two nodes of which at least one is in C over every node in C.*

Proof. Assume the contrary. This means there exist a node $X \in C$ such that no node in C forms an unshielded collider over X. This does however mean there exists an essential MVR CG G' with the same structure as G^* with the exception that X is a parent of C instead of in it. This in turn means that G' and G^* represents the same Markov equivalence class but that G' contains less arrowheads than G^* which is a contradiction. □

Lemma 7. *Let G^* and G' be two essential MVR CGs without any undirected edges such that G^* and G' have the same nodes and structure with the exception that G' is missing a component C that exists in G^* such that $ch_{G^*}(C) = \emptyset$. The operators in Definition 2 can then transform G' to G^* through a sequence of essential MVR CG $G_1, \ldots, G_n$ by only applying one operator at a time.*

Proof. If $|C| = 1$ this follows from Lemma 8, otherwise it follows from Lemma 9. □

Let G^* and G' be two essential MVR CGs without any undirected edges such that G^* and G' have the same nodes and structure with the exception that G' is missing a component C containing only one node Y, that exists in G^* such that $ch_{G^*}(Y) = \emptyset$. The algorithm below then defines the operations and their order to transform G' into G^* through a sequence of essential MVR CGs $G_1, \ldots, G_n$. Let G_i be the input graph for each line that is transformed into G_{i+1}, which then takes the place of G_i in the next executed line:

1 Add Y to G' and denote the new graph G_1
2 Repeat until no case is applicable (restart the loop after each change):
3 If there exist a node $X \in pa_{G^*}(Y)$ such that $X \notin pa_{G_i}(Y)$ and $sp_{G^*}(X) \neq \emptyset$ then let $G_{i+1} = G_i \cup \{X \to Y\}$.
4 If there exist a node $X \in pa_{G^*}(Y)$ such that $X \notin pa_{G_i}(Y)$ and $bd_{G^*}(X) \not\subseteq pa_{G^*}(Y)$ then let $G_{i+1} = G_i \cup \{X \to Y\}$.
5 If there exist a node $X \in pa_{G^*}(Y)$ such that $X \notin pa_{G_i}(Y)$ and $\exists Z \in pa_{G_i}(Y) \cap (co_{G_i}(X) \cup de_{G_i}(X))$ then let $G_{i+1} = G_i \cup \{X \to Y\}$.
6 If there exist a node $X \in pa_{G^*}(Y)$ such that $X \notin pa_{G_i}(Y)$ and $pa_{G_i}(Y) \not\subseteq bd_{G^*}(X)$ then let $G_{i+1} = G_i \cup \{X \to Y\}$.
7 If there exist two nodes $X, Z \in pa_{G^*}(Y)$ such that $X, Z \notin pa_{G_i}(Y)$ and $X \notin ad_{G^*}(Z)$ then let $G_{i+1} = G_i \cup \{X \to Y, Z \to Y\}$.

Algorithm 2: Procedure for adding edges when $|C| = 1$

Lemma 8. *All graphs $G_1, \ldots, G_n$ described in Algorithm 2 are essential MVR CGs and the transformation in each step can be achieved through one or more operations described in Definition 2 in which case all intermediate graphs are essential MVR CGs.*

Proof. Assume the contrary. This means that one of the following statements hold: (1) One of the lines in Algorithm 2 creates a graph G_{i+1} that is not an essential MVR CG when G_i is an essential MVR CG. (2) That G_i is not transformable into G_{i+1} through a sequence of operations described in Definition 2 or (3) that $G_i \neq G^*$ holds for the essential MVR CG G_i that passes line 7. We will first show that (1) and (2) must be false and then finish the proof contradicting that (3) can hold. Note that since G' is an essential MVR CG we know that all edges in the subgraph of G_i induced by $V_{G_i} \setminus Y$ must be indifferent. This follows from the fact that the algorithm only add directed edges oriented towards Y and hence that the cause making any arrowhead indifferent in G' must still be valid in G_i. Also note that G_i for any i must be a subgraph of G^* and hence that $ad_{G_i}(Y) = pa_{G_i}(Y)\ \forall i$.

We will first study G_{i+1} for all lines in algorithm 2. To see that G_{i+1} cannot contain any semi-directed cycle it is enough to note that G_{i+1} is a subgraph of G^* which can not contain any semi-directed cycles. Furthermore any new arrowhead on a new edge in G_{i+1} (compared to G_i) must be and remain indifferent in G_j for all $j \geq i$ for the following reasons: line 3, since a directed edge that share an endnode with a bidirected edge always is indifferent according to Lemma 4. Line 4, since any other orientation of the edge would result in an unshielded collider not in G^*. Line 5, since any other orientation of the edge would cause a semi-directed cycle to appear in G_j, $j > i$. To see this note that all arrowheads in G_i must be remain indifferent in all G_j $j \geq i$. This can be seen inductively starting with G_1 which have the same edges as $G_i \setminus Y$ which must be indifferent for all G_i as noted above. For each iteration i of the algorithm one or more new directed edges is then added with the property that they will remain indifferent for all $j \geq i$. Hence all arrowheads in G_i must remain indifferent for all $j \geq i$ and hence any other orientation of the edge added in line 5 must cause a semi-directed cycle in all G_j $j \geq i$. Line 6, since an unshielded collider that exists in G^* is added and line 7 since $X \to Y$ and $Z \to Y$ both are part of an unshielded collider in G^*. Hence G_{i+1} must be essential MVR CGs for these lines.

In line 7 we will however have to do two or three operations to reach G_{i+1} from G_i and hence we have to show that all intermediate graphs are essential MVR CGs. Note that if this line is reached then (1) $bd_{G_i}(Y) \subseteq bd_{G_i}(X)$ (resp. $bd_{G_i}(Y) \subseteq bd_{G_i}(Z)$) and $sp_{G_i}(X) = sp_{G_i}(Z) = sp_{G_i}(Y) = \emptyset$ must hold. (1) must hold or line 6 must be applicable since $ad_{G_i}(X) \setminus Y = ad_{G^*}(X) \setminus Y$ (resp. $ad_{G_i}(Z) \setminus Y = ad_{G^*}(Z) \setminus Y$) and (2) must hold since we know that $sp_{G^*}(Y) = \emptyset$ and if $sp_{G_i}(X) \neq \emptyset$ (resp. $sp_{G_i}(Z) \neq \emptyset$) line 3 must be applicable.

We can now have two cases. If $bd_{G_i}(X) \not\subseteq bd_{G_i}(Y)$ (resp. $bd_{G_i}(Z) \not\subseteq bd_{G_i}(Y)$) then we know that $G_i \cup \{X \to Y\}$ (resp. $G_i \cup \{Z \to Y\}$) must be an essential MVR CG. Hence we can first add $X \to Y$ (resp. $Z \to Y$) to G_i whereafter we can add $Z \to Y$ (resp. $X \to Y$) and then reach G_{i+1}. However if both $bd_{G_i}(X) \subseteq bd_{G_i}(Y)$ and $bd_{G_i}(Z) \subseteq bd_{G_i}(Y)$ hold then we know that $bd_{G_i}(X) = bd_{G_i}(Z) = bd_{G_i}(Y)$ must hold. Hence both $G_i \cup \{X - Y\}$ and $G_i \cup \{X - Y, Y - Y\}$ must be essential MVR CGs. We can then perform the "add V-collider" operation to transform the undirected edges to directed edges oriented towards Y in $G_i \cup \{X - Y, Y - Y\}$

to reach G_{i+1}. Hence all intermediate graphs for line 7 must be essential MVR CGs.

The last thing to prove is then that $pa_{G_i}(Y) = pa_{G^*}(Y)$ after line 7. To see this assume the contrary, i.e. that there exist a node $X \in pa_{G^*}(Y)$ but $X \notin pa_{G_i}(Y)$. In addition, let us choose this X such that no node W exists for which $W \in (de_{G^*}(X) \setminus co_{G^*}(X)) \cap pa_{G^*}(Y)$ and $W \notin pa_{G_i}(Y)$ hold. Obviously $sp_{G_i}(X) = \emptyset$ or line 3 would be applicable. We also know that $pa_{G^*}(X) = pa_{G_i}(X) \subseteq pa_{G^*}(Y)$ or line 4 would be applicable. $bd_{G^*}(Y) \subseteq ad_{G^*}(X)$ must also hold since no unshielded collider can exist between X and some node Z over Y in G^* or line 6 (if $Z \in pa_{G_i}(Y)$) or line 7 (if $Z \notin pa_{G_i}(Y)$) would be applicable. This means that $pa_{G^*}(X) = pa_{G^*}(Y) \setminus X$. Finally we also know that reversing the orientation of the edge $X \to Y$ in G_i would cause no semi-directed cycle to appear or line 5 would be applicable. This does however, since no node W existed and hence that all descendants of X already is added as parents to Y, that the orientation of the edge $X \to Y$ also can be reversed in G^* without creating any semi-directed cycle or unshielded collider. This does however contradict that G^* is an essential MVR CG. Hence we have a contradiction that X can exists after line 7. □

Let G^* and G' be two essential MVR CGs without any undirected edges such that G^* and G' have the same nodes and structure with the exception that G' is missing a component C, such that $|C| \geq 2$, that exists in G^* such that $ch_{G^*}(C) = \emptyset$. The algorithm below then defines the operations and their order to transform G' into G^* through a sequence of essential MVR CGs $G_1, \ldots, G_n$. Let G_i be the input graph for each line that is transformed into G_{i+1}, which then takes the place of G_i in the next executed line:

1 Add all the nodes in C to G' and denote the new graph G_1
2 Repeat until until $G_i = G^*$:
3 Let X, Y and Z be three nodes in G_i such that $X \in C, Y \in C$, $Z \in pa_{G^*}(C) \cup C$, X and Z form an unshielded collider over Y in G^*, $bd_{G_i}(Y) = \emptyset$ and $\forall W \in ch_{G_i}(Y)$ such that $C_w = co_{G_i}(W)$ there $\exists R, P \in C_w$ such that $bd_{G_i}(R) \not\subseteq bd_{G^*}(Y) \cup Y$, $P \in ch_{G_i}(Y)$ and $bd_{G_{i+1}}(Y) \not\subseteq bd_{G^*}(P) \cup P$ hold. Let $G_{i+1} = G_i$ with the exception that all non-arrowheads towards Y are replaced by arrowheads and if X and/or Z is not adjacent of Y in G_i then $X \to Y$ and/or $Z \to Y$ is added to G_{i+1}.
4 Repeat until $bd_{G_i}(Y) = bd_{G^*}(Y)$:
5 Let Q be a node such that $Q \in bd_{G^*}(Y)$ but $Q \notin bd_{G_i}(Y)$. Then let $G_{i+1} = G_i \cup \{Q \to Y\}$.

Algorithm 3: Procedure for adding edges when $|C| > 1$

Lemma 9. *All graphs $G_1, \ldots, G_n$ described in Algorithm 3 are essential MVR CGs and the transformation in each step can be achieved through one or more operations described in Definition 2 in which case all intermediate graphs are essential MVR CGs.*

Proof. Assume the contrary. This means that one of the following statements hold: (1) One of the lines in Algorithm 3 creates a graph G_{i+1} that is not an essential MVR CG when G_i is an essential MVR CG. (2) That G_i is not transformable into G_{i+1} through a set of operations described in Definition 2 or (3) that G_i never becomes G^*. We also make two assumptions that have to be proven. Assumption 1 is that when when the loop in line 2 restarts we have that $\forall c_j \in C$, such that $bd_{G_i}(c_j) \neq \emptyset$, $bd_{G_i}(c_j) = bd_{G^*}(c_j)$ must hold for the current graph G_i. We will denote these c_j as being "collided". Assumption 2 is that all bidirected edges in any G_i must be indifferent in all G_j $j \geq i$.

The rest of the proof is constructed as follows. In part 1 we will first show that assumption 1 holds. This is then followed by proving that all G_{i+1} are essential MVR CGs for line 3 in part 2. In the end of part 2 we also prove assumption 2. In part 3 we then show that G_i is transformable into G_{i+1} through a set of operations described in Definition 2 and that all intermediate graphs also are essential MVR CGs. In part 4 we show that G_{i+1} must be an essential MVR CG for line 5 and finally in part 5 we show that the algorithm must terminate and hence that G_i becomes G^* after $|C|$ number of iterations of line 2.

First we will however make some observations about the algorithm and G_i. Note that only two lines, line 3 and line 5, changes the structure of G_i to G_{i+1}. For these lines we can see that a directed edge $X \to Y$ only is added to G_{i+1} if $X \to Y$ or $X \leftrightarrow Y$ is in G^*. We can also see that a bidirected edge $X \leftrightarrow Y$ only is added to G_{i+1} in line 3 and then only if it also exists in G^* since any $X \in ch_{G_i}(Y)$ and Y must both be adjacent and in the same bidirected component in G^*. We can also note that edges never are removed once they have been added to G_i, although directed edges can be replaced by bidirected edges. This means that once an arrowhead is added to G_{i+1} it must exist in all G_j $j > i$. Finally we can also note that edges only are added with arrowheads towards collided nodes, i.e. nodes that has been chosen as Y in line 3.

Part 1:
That $\forall c_j \in C$, such that $bd_{G_i}(c_j) \neq \emptyset$, $bd_{G_i}(c_j) = bd_{G^*}(c_j)$ obviously holds for $i = 1$ since $bd_{G_1}(c_j) = \emptyset$ $\forall c_j \in C$. Then, for each iteration of line 2, one node Y is selected in line 3 and becomes collided. When a node is chosen as Y in line 3 two things happen; First, in line 3, all already adjacent nodes of Y become spouses of Y when all non-arrowheads towards Y are replaced with arrowheads. Secondly, in line 5, that directed edges are added from all remaining nodes in $bd_{G^*}(Y)$ to Y. From this it is clear that once this is done all nodes in $bd_{G^*}(Y)$ also must be in $bd_{G_i}(Y)$.

We can also note that the algorithm never adds any edges oriented towards any other nodes than those chosen as Y, nor does it ever remove any arrowheads. This means that all nodes $c_j \in C$, for which $bd_{G_i}(c_j) \neq \emptyset$ holds, must have been chosen as Y in some iteration k, and hence we know that $bd_{G_l}(c_j) = bd_{G^*}(c_j)$ must hold $\forall l > k$. Hence it is clear from induction, starting with $i = 1$ that $\forall c_j \in C$, such that $bd_{G_i}(c_j) \neq \emptyset$, $bd_{G_i}(c_j) = bd_{G^*}(c_j)$ must hold $\forall G_i$ hence that assumption 1 holds.

Part 2:
In this part we study line 3 and show that all G_{i+1} are essential MVR CGs for line 3 and in the end we also show that assumption 2 holds. Here we have three nodes $X \in C$, $Y \in C$ and $Z \in pa_{G^*}(C) \cup C$ such that X and Z forms an unshielded collider over Y in G^*. We also know that $bd_{G_i}(Y) = \emptyset$ and $\forall W \in ch_{G_i}(Y)$ such that $C_w = co_{G_i}(W)$ there $\exists R, P \in C_w$ such that $bd_{G_i}(R) \not\subseteq bd_{G^*}(Y) \cup Y$, $P \in ch_{G_i}(Y)$ and $bd_{G_{i+1}}(Y) \not\subseteq bd_{G^*}(P) \cup P$ hold.

We can also make some deductions about the structure of G_i. Recall that $ch_{G^*}(C) = \emptyset$ and hence we know, together with assumption 1, that no collided nodes in C can have any children. From assumption 1 we also know that any child of Y, as well as any node in C that is endnode of a bidirected edge, must be collided. This in turn means that $\forall c_j \in C$ such that $bd_{G_i}(c_j) \neq \emptyset$, we must have that $ch_{G_i}(c_j) = \emptyset$ which means that $de_{G_i}(c_j) \setminus co_{G_i}(c_j) = \emptyset$ must hold.

Also note that we can have three subcases: Case 1, that both $X, Z \in ch_{G_i}(Y)$, case 2, that $X \in ch_{G_i}(Y)$ but $Z \notin ad_{G_i}(Y)$ and finally case 3, that $X, Z \notin ad_{G_i}(Y)$. If we have case 1 line 3 then consists of replacing all non-arrowhead edgeendings towards Y with an arrowheads simultaneously. Hence all children of Y will become spouses of Y. In case 2 all non-arrowheads are replaced for G_{i+1} but $Z \to Y$ is also added. Finally in case 3 both $X \to Y$ and $Z \to Y$ is added to G_{i+1} in addition to the replacement of the non-arrowheads.

We can now prove that G_{i+1} cannot contain any semi-directed cycle since we know that $bd_{G_i}(Y) = \emptyset$ and that $\forall c_j \in ch_{G_i}(Y)$ we must have that $de_{G_i}(c_j) \setminus co_{G_i}(c_j) = \emptyset$. If we assume case 1 we know that no new directed edges are added and hence that $\forall c_j \in ch_{G_i}(Y)$ $ch_{G_i}(c_j) = ch_{G_{i+1}}(c_j)$. We also know that $ch_{G_{i+1}}(Y) = \emptyset$ and, since $\forall c_j \in co_{G_{i+1}}(Y) \cup Y$ $ch_{G_{i+1}}(c_j) = \emptyset$, we obviously cannot have a semi-directed cycle. For case 2 we can note that we can have two cases, either $Z \in C$ or $Z \notin C$. If $Z \in C$ then we know that Z cannot be previously collided or $Z \in ch_{G_i}(Y)$ would have to hold due to assumption 1 and that $Z \in sp_{G^*}(Y)$. Hence we know that $bd_{G_i}(Z) = \emptyset$ and since the only edge containing Z as an endnode that is added in line 3 is $Z \to Y$ and hence $bd_{G_{i+1}}(Z) = \emptyset$ must hold. This does however contradict that Z is part of any semi-directed cycle. If we on the other hand have that $Z \notin C$ then we know that $\forall c_j \in C$ $Z \notin de_{G_i}(c_j)$ and $Z \notin de_{G_{i+1}}(c_j)$ must hold. Hence we cannot have a semi-directed cycle containing Z in G_{i+1} for case 2 and that no semi-directed cycle can exist not containing Z follows similarly as for case 1. For case 3 we can see that no semi-directed cycle can exist in G_{i+1} and contain Z or X similarly as we saw in case 2 since X must be in C. It also follows similarly as for case 1 that no semi-directed cycle can exist in G_{i+1} containing neither X nor Z. Hence no semi-directed cycle can exist in G_{i+1}.

Now assume that a not indifferent arrowhead is created in G_{i+1} when the non-arrowheads are replaced by arrowheads. We know that the edge $X \to Y$ (resp. $Z \to Y$), for case 2 or 3 (resp. case 3) must be part of an unshielded collider over Y and hence be indifferent. This means that the not indifferent arrowhead must be on one of the bidirected edges created containing Y as an endnode. Let S be the node for which either the arrowhead towards Y or towards

S, or both, on the edge $S \leftrightarrow Y$ is not indifferent in G_{i+1}. We can then note that S, as well as all nodes in $C_S = co_{G_i}(S)$, must be collided previously since $Y \rightarrow S$ is in G_i due to assumption 1. From the prerequisite of choosing Y in line 3 we know that there $\exists c_n, d_m \in C_S$ such that $bd_{G_i}(c_n) \not\subseteq bd_{G^*}(Y) \cup Y$, $d_m \in ch_{G_i}(Y)$ and $bd_{G_{i+1}}(Y) \not\subseteq bd_{G^*}(d_m) \cup d_m$ hold. Note that c_n and d_m can be S. Let $S \leftrightarrow c_1 \leftrightarrow c_2 \leftrightarrow \ldots \leftrightarrow c_n$, $n \geq 0$ (0 if $c_n = S$), be a path π_1 between S and c_n in G_i. Without losing generalization we can assume that c_n must be adjacent of Y or there must exist some node $c_k \in S \cup c_1 \cup \ldots \cup c_{n-1}$ such that c_k is adjacent of Y and such that $bd_{G_i}(c_k) \not\subseteq bd_{G^*}(Y) \cup Y$ and hence can take the place of c_n. Hence we know that $Y \leftrightarrow c_n$ must be in G_{i+1}. From assumption 2 we then know that all bidirected edges on π_1 must be indifferent in G_{i+1} since they also existed in G_i. From $bd_{G_i}(c_n) \not\subseteq bd_{G^*}(Y) \cup Y$ we then know that the arrowhead towards c_n on $Y \leftrightarrow c_n$ must be indifferent in G_{i+1} since it is part of an unshielded collider in G_i and $bd_{G_{i+1}}(c_n) = bd_{G_i}(c_n)$. This in turn means that the arrowhead towards S on $Y \leftrightarrow S$ must be indifferent or a semi-directed cycle would exist in G_{i+1}. Similarly we can see that there must exist a path π_2 of indifferent bidirected edges between S and d_m and that the arrowhead towards Y on $Y \leftrightarrow d_m$ must be indifferent in G_{i+1}. This is due to that $bd_{G_{i+1}}(Y) \not\subseteq bd_{G^*}(d_m) \cup d_m$ holds according to the prerequisite and $bd_{G^*}(d_m) = bd_{G_i}(d_m)$ since d_m has been collided previously and hence that d_m forms an unshielded collider over Y with some other node in G_{i+1}. This in turn means that the arrowhead towards Y on $Y \leftrightarrow S$ must be indifferent or a semi-directed cycle occurs in G_{i+1}. Hence we have shown that the arrowhead towards S is indifferent due to c_n and that the arrowhead towards Y is indifferent due to d_m on the edge $S \leftrightarrow Y$ and contradicted that any arrowhead on this edge can be indifferent. Hence all arrowheads added in line 3 must be indifferent.

Furthermore we can also see that the reasoning must hold for any edge $S \leftrightarrow Y$ for all G_j $j > i$ because of the following: (1) the arrowhead towards c_n on $c_n \leftarrow \rightarrow Y$ must be indifferent in all G_j since $bd_{G_i}(c_n) = bd_{G^*}(c_n)$ must hold since c_n has been collided previously. This then means that $bd_{G_j}(c_n) \not\subseteq bd_{G^*}(Y) \cup Y$ must hold for all $j > i$. (2) the arrowhead towards Y on $d_m \leftrightarrow Y$ must be indifferent in all G_j since $bd_{G_{i+1}}(Y) \subseteq bd_{G_j}(Y)$ and hence $bd_{G_j}(Y) \not\subseteq bd_{G^*}(d_m) \cup d_m$ must hold for all j. (3) Finally we can show that all bidirected edges in $co_{G_i}(S)$ must be indifferent for all G_j $j > i$ iteratively. Starting with G_1 containing no edges with endnodes in C this statement obviously holds, and each time a bidirected edge is added to G_i it must hold that all previously added bidirected edges in G_i must be indifferent for all G_j $j \geq i$. Therefore, following the reasoning above, the bidirected edges between any nodes in $sp_{G_{i+1}}(Y)$ and Y must also be indifferent for all G_j $j > i$ and we have proven that assumption 2 is correct since this is the only place in the algorithm where bidirected edges are created.

What remains to prove to show that G_{i+1} is an essential MVR CG is that no arrowhead are made not indifferent in line 3. Due to assumption 2 we then know that the edge made not indifferent cannot be a bidirected edge, and since G_i contains no undirected edges, we know that it must be a directed edge. For a directed edge to have an indifferent arrowhead it must either be part of an

unshielded collider, or the other orientation of it must cause a new unshielded collider or a semi-directed cycle to occur in the graph. Hence, since edges only are added and made bidirected in line 3, for an edge to be made not indifferent it must cease to be part of an unshielded collider. This means that we can, without losing generality, assume a directed edge with an endnode M cease to be part of an unshielded collider when $X \to Y$ is added to G_i. Then we know that $X \to M$ and $Y \to M$ also must exist in G_i since X and Y must form an unshielded collider over M in G_i but neither X nor Y have been collided previously. This does however mean that $Y \leftrightarrow M$ must exist in G_{i+1} which together with Lemma 4 contradicts that any adjacent directed edge can be not indifferent. Hence we have a contradiction and all arrowheads in G_{i+1} must be indifferent which means that G_{i+1} is an essential MVR CG for line 3.

Part 3
It is easy to see that G_i is transformable into G_{i+1} in line 5 since the transformation only consists of adding a directed edge and hence that operation can be used. For line 3 it is however more difficult. In part 2 three subcases, case 1, 2 and 3, were identified. If we have case 1 then it is easy to see that the transformation from G_i to G_{i+1} can be achieved by the "add V-collider" operation.

For case 2 we can note three things; first that $bd_{G_i}(Y) = \emptyset$, secondly that X is collided and hence that $Y \to X$ is in G_i. Thirdly we can note that we can have two major sub-cases, either $bd_{G_i}(Z) = \emptyset$ or $bd_{G_i}(Z) \neq \emptyset$. First assume that $bd_{G_i}(Z) = \emptyset$ holds. We can then add an undirected edge between Z and Y. The resulting graph $G_i \cup \{Z{-}Y\}$ must be an essential MVR CGs since $bd_{G_i}(Y) = bd_{G_i}(Z) = \emptyset$ and since G_i did not contain any undirected edges the created undirected component is chordal. To see that no indifferent arrowheads are made not indifferent assume the contrary. Then there must exist a third node M such that $Y \to M$ and $Z \to M$ are in G_i for which either the arrowhead on $Z \to M$ or $Y \to M$ is made indifferent similarly as we saw in part 2. From assumption 1 we know that $bd_{G_i}(M) = bd_{G^*}(M)$ and from Lemma 6 we know that there must exist an unshielded collider over M. Hence we know that an unshielded collider exists over M in G_i. Note that Z and Y cannot be the two nodes forming the unshielded collider over M since Z is adjacent of Y in G^*. Let U be one of the nodes in $bd_{G^*}(M) \setminus (Z \cup Y)$ forming such an unshielded collider. If $U \in pa_{G^*}(C)$ or if $U \in C$ but has not been collided previously then there can exist no edge between U and Y since $bd_{G_i}(Y) = \emptyset$ and $bd_{G_i}(U) \cap C = \emptyset$. Hence $Y \to M$ must form an unshielded collider with U over M in G_{i+1} which gives us a contradiction. If $U \in C$ and U has been collided we know that $U \leftrightarrow M$ must be in G_i and from Lemma 4 we then have that any directed edge sharing an endnode M with an indifferent bidirected edge must be indifferent. Hence we have a contradiction and $G_i \cup \{Z{-}Y\}$ must be an essential MVR CG. From $G_i \cup \{Z{-}Y\}$ we can then perform the "add V-collider" operation to reach G_{i+1}.

If on the other hand $bd_{G_i}(Z) \neq \emptyset$ we know that Z cannot be in C or $Y \to Z$ would exist in G_i according to assumption 1. We also know that $G_i \cup \{Z \to Y\}$ must be an essential MVR CG since reversing the edge would cause the graph to have a different set of unshielded colliders. To see this note that $bd_{G_i}(Y) = \emptyset$

and that $nb_{G_i}(Z) = \emptyset$. Hence the edge $Y \to Z$ would cause Y to form unshielded colliders over Z with all nodes in $bd_{G_i}(Z)$ and hence the edge $Z \to Y$ must be indifferent. Also note that adding $Z \to Y$ cannot cause some already existing arrowhead in G_i to become not indifferent according to the same reasoning as for $Z-Y$ above. Hence $G_i \cup \{Z \to Y\}$ must be an essential MVR CG.

For case 3 we can follow the same reasoning. In this case we do however know that X cannot have been collided previously and hence that $bd_{G_i}(X) = \emptyset$ must hold due to assumption 1. First assume that $bd_{G_i}(Z) = \emptyset$ holds. We can then add undirected edges, first between X and Y and then between Z and Y. The resulting graphs $G_i \cup \{X-Y\}$ and $G_i \cup \{X-Y, Z-Y\}$ must be essential MVR CGs since $bd_{G_i}(X) = bd_{G_i}(Y) = bd_{G_i}(Z) = \emptyset$ and since G_i did not contain any undirected edges the created undirected component must be chordal. That no indifferent arrowheads are made not indifferent follows as for case 2 above. From $G_i \cup \{Z-Y, X-Y\}$ the "add V-collider" operation can then be performed to reach G_{i+1}.

If on the other hand $bd_{G_i}(Z) \neq \emptyset$ it follows similarly as for case 2 that $G_i \cup \{Z \to Y\}$ is an essential MVR CG. From this graph the edge $X \to Y$ can then be added which now will form an unshielded collider, since $bd_{G_i}(X) = \emptyset$, with Z over Y and hence $G_i \cup \{Z \to Y, X \to Y\}$ must be an essential MVR CG. Note that adding $Z \to Y$ (resp. $X \to Y$) cannot cause some already existing arrowhead in G_i to become not indifferent according to the same reasoning as for $X-Y$ above.

This gives us that G_i is transformable into G_{i+1} using the operators in Definition 2 for all lines in Algorithm 3 and that all intermediate graphs are essential MVR CGs.

Part 4

In this part we will show that G_{i+1} in line 5 must be an essential MVR CG if G_i is an essential MVR CG for the same line. Line 5 add a directed edge $Q \to Y$ to G_i for which we know that $\forall c_k \in co_{G_i}(Y)$ c_k must be collided, due to assumption 1, and hence that $ch_{G_i}(c_k) = \emptyset$ similarly as we saw in part 2. Hence we know that $de_{G_i}(Y) \setminus co_{G_i}(Y) = de_{G_{i+1}}(Y) \setminus co_{G_{i+1}}(Y) = \emptyset$. From this it directly follows that G_{i+1} can contain no semi-directed cycle since G_i contains no semi-directed cycle.

Now assume that the arrowhead on the edge $Q \to Y$ is not indifferent in G_{i+1}. We know there exists a node X as described in line 3 in the algorithm such that $X \in bd_{G_i}(Y) \cap C$. Since $Y \leftrightarrow Q$ is not in G_i we must also have that either $Q \notin C$ holds or that Q has not been collided previously. In either case we know that $Q \notin ch_{G_i}(X)$ must hold. First assume that X has not been collided previously. Then, since $bd_{G_i}(X) = \emptyset$, X and Q cannot be adjacent and therefore must form an unshielded collider over Y and hence the arrowhead on the edge $Q \to Y$ must be indifferent in G_{i+1}. If X on the other hand has been collided previously we know that both G_i and G_{i+1} must contain the bidirected edge $X \leftrightarrow Y$. From Lemma 4 we then get that any directed edge with X or Y as an endnode also must be indifferent and hence we have a contradiction that $Q \to Y$ is not indifferent in G_{i+1}.

Secondly assume that the edge $Q \rightarrow Y$ causes some other previously indifferent arrowhead to become not indifferent in G_{i+1}. Similarly as we saw in the last paragraph of part 2 we know that this must include a third node $M \in ad_{G_i}(Q) \cap ad_{G_i}(Y)$ such that there is either the edge between Q and M or Y and M that is made not indifferent. We can from this note that the edge between Y and M cannot be bidirected since Lemma 4, together with assumption 2, then states that all directed edges containing M as an endnode must be indifferent. Hence either $Y \rightarrow M$ or $M \rightarrow Y$ must exist in G_i. However, since only directed edges oriented towards Y is added after line 3, and every child of Y is made a spouse of Y in line 3, we can see that the edge $Y \rightarrow M$ cannot exist in G_i in line 5. Hence $M \rightarrow Y$ must exist in G_i. From this it follows that, since Y and X cannot form an unshielded collider over M in G_i when $M \rightarrow Y$ is in G_i, it must be the edge $M \rightarrow Y$ that is made not indifferent when $Q \rightarrow Y$ is added. However, just like in the last paragraph we can note that a node X must exist, and that if X is collided then $M \rightarrow Y$ must be indifferent due to the bidirected edge, while if it is not collided, then M and X must form an unshielded collider over Y in G_{i+1}. Hence the edge $M \rightarrow Y$ must be indifferent in G_{i+1} and we have a contradiction which means that G_{i+1} for line 5 must be an essential MVR CG if G_i is an essential MVR CG.

Part 5
The last thing to prove is that the algorithm must terminate and hence that G_i becomes G^* after $|C|$ number of iterations of line 2.

To show this assume the contrary, i.e. that there exist no node in G_i for which line 3 is applicable but where there exist a node Y that has not yet been collided and hence $bd_{G_i}(Y) \neq bd_{G^*}(Y)$. From Lemma 6 we know that an unshielded collider must exist over Y in G^* for which at least one of the boundary nodes are in C. From assumption 1 we also know that $\forall c_j \in C$, such that $bd_{G_i}(c_j) \neq \emptyset$, $bd_{G_i}(c_j) = bd_{G^*}(c_j)$ must hold for G_i and hence that $bd_{G_i}(Y) = \emptyset$ must hold. It is now easy to see that $ch_{G_i}(Y) \neq \emptyset$ must hold or line 3 would be applicable. In general this also means that for all non-collided nodes $c_l \in C$ we must have that $ch_{G_i}(c_l) \neq \emptyset$ or line 3 would be applicable for that node.

Before we continue we will note two things (1) First that for every connected set of nodes $A \subset C$, i.e. every $A \subset C$ for which there exist a path between any pair of nodes in the subgraph of G^* induced by A, we must have that there $\exists a_l \in A$ such that $\exists R \in sp_{G^*}(a_l)$ and $bd_{G^*}(a_l) \not\subseteq bd_{G^*}(R) \cup R$. If this would not be the case there would exist no unshielded collider over A. Hence there would exist a MVR CG G', such that $I(G') = I(G^*)$, where G' has the same structure as G^* with the exception that A is a parent component of $C \setminus A$ in G' instead of in the same component as $C \setminus A$ as in G^*. This is of course a contradiction since G' would contain less arrowheads than G^* and hence G^* cannot be an essential MVR CG. Similarly we can note (2) that $\exists a_m \in A$ such that $\exists P \in sp_{G^*}(a_m)$ and $bd_{G^*}(P) \not\subseteq bd_{G^*}(a_m) \cup a_m$ or A could be a child component of $C \setminus A$ and the graph would still represent the same independence model.

Let $B \subset C$ such that B forms a component in G_i. Obviously such a B must exist or C would form one large component and hence $G_i = G^*$. From (2) we

know there exists at least one node $b_r \in B$ such that there $\exists T \in pa_{G_i}(b_r) \cap C$ such that $bd_{G^*}(T) \not\subseteq bd_{G^*}(b_r) \cup b_r$. To see this let A in (2) take the form of B. Then we know there exists a node $a_m \in A$ such that $\exists P \in sp_{G^*}(a_m)$ such that $bd_{G^*}(P) \not\subseteq bd_{G^*}(a_m) \cup a_m$. Obviously $P \in sp_{G_i}(a_m)$ cannot hold, since P then would belong to B, but since we know that $\forall b_k \in B\ bd_{G_i}(b_k) = bd_{G^*}(b_k)$ by assumption 1, and that $P \in bd_{G^*}(a_m)$, we can deduce that $P \in pa_{G_i}(a_m) \cup nb_{G_i}(a_m)$ must hold. This together with the fact that G_i does not contain any undirected edges means that $P \in pa_{G_i}(a_m)$. Hence there must exist a b_r taking the form of a_m and T taking the form of P. Also note that if different nodes in $ch_{G_i}(T)$ belong to different components in G_i, then there must exist a node with the same properties as b_r in all components. To see this note that $ch_{G_i}(T) \subseteq sp_{G^*}(T)$ and that all nodes in $ch_{G_i}(T)$ must have been collided according to assumption 1. For the prerequisites of line 3 not to be fulfilled we know that for some component D that contains a child of T we must have that $\forall d_k \in D$ $bd_{G_i}(d_k) \subseteq bd_{G^*}(T) \cup T$ must hold. Moreover, since d_k must be collided we must also have that $bd_{G_i}(d_k) = bd_{G^*}(d_k)$ and hence that $bd_{G^*}(d_k) \subseteq bd_{G^*}(T) \cup T$ must hold. This in turn also means that $\forall d_k \in D\ d_k \in ch_{G_i}(T)$ must hold. If we let $E = pa_{G_i}(D) \cap C$ and F consist of all nodes $e_m \in E$ such that $\exists d_n \in D$ and $bd_{G^*}(e_m) \not\subseteq bd_{G^*}(d_n) \cup d_n$ we can note that $\forall f_m \in F$ we must have that $\forall d_k \in D\ bd_{G_i}(d_k) \subseteq bd_{G^*}(f_m) \cup f_m$ and $d_k \in ch_{G_i}(f_m)$ or the prerequisite for line 3 must be fulfilled for a node f_m. In addition, since d_k must be collided in G_i we must have that $bd_{G^*}(d_k) \subseteq bd_{G^*}(f_m) \cup f_m$ holds. This does however mean that F is complete and that $\forall f_m \in F\ D \cup (E \setminus F) \subseteq ad_{G^*}(f_m)$ and $ad_{G^*}(D \cup (E \setminus F)) \subseteq ad_{G^*}(f_m) \cup f_m$. This means that there exists no colliders towards any node in $D \cup (E \setminus F)$ from any other node in C. Hence we have a contradiction with (1) since there then would exist an essential MVR CG G' with the same structure as G^* with the exception that $D \cup (E \setminus F)$ are parents of C instead of in C which contradicts that G^* is an essential MVR CG. Hence at least one of the nodes in F must be possible to collide which contradicts the assumption. □

Acknowledgments This work is funded by the Center for Industrial Information Technology (CENIIT) and a so-called career contract at Linköping University, by the Swedish Research Council (ref. 2010-4808), and by FEDER funds and the Spanish Government (MICINN) through the projects TIN2010-20900-C04-03 and TIN2010-20900-C04-01.

Bibliography

[1] Ali, R.A., Richardson, T.S., Spirtes, P., Zhang, J.: Towards Characterizing Markov Equivalence Classes for Directed Acyclic Graphs with Latent Variables. In: Proceedings for the 21st Conference on Uncertainty in Artificial Intelligence. pp. 10–17 (2005)

[2] Andersson, S.A., Madigan, D., Pearlman, M.D.: An Alternative Markov property for Chain Graphs. Scandianavian Journal of Statistics 28, 33–85 (2001)

[3] Andersson, S.A., Madigan, D., Perlman, M.D.: A characterization of Markov Equivalence Classes for Acyclic Digraphs. Annals of Statistics 25:2, 505–541 (1997)

[4] Andersson, S.A., Madigan, D., Perlman, M.D.: On the Markov Equivalence of Chain Graphs, Undirected Graphs and Acyclic Digraphs. Scandianavian Journal of Statistics 24, 81–102 (1997)

[5] Cox, D.R., Wermuth, N.: Linear Dependencies Represented by Chain Graphs. Statistical Science 8, 204–283 (1993)

[6] Frydenberg, M.: The Chain Graph Markov Property. Scandinavian Journal of Statistics 17, 333–353 (1990)

[7] Gillispie, S.B., Perlman, M.: The Size Distribution for Markov Equivalence Classes of Acyclic Digraph Models. Artificial Intelligence pp. 137–155 (2002)

[8] Goudie, R.J.B., Mukherjee, S., Griffiths, F.: Exploratory Network Analysis of Large Social Science Questionnaires. In: Proceedings of the 8th UAI Bayesian Modeling Applications Workshop. pp. 34–42 (2011)

[9] Häggström, O.: Finite Markov Chains and Algorithmic Applications. Cambridge University Press (2002)

[10] Harary, F., Palmer, E.M.: Graphical Enumeration. Academic Press (1973)

[11] He, Y., Jia, J., Yu, B.: Reversible MCMC on Markov Equivalence Classes of Sparse Directed Acyclic Graphs. The Annals of Statistics pp. 1742–1779 (2013)

[12] Koller, D., Friedman, N.: Probabilistic Graphcal Models. MIT Press (1999)

[13] Lauritzen, S.L., Wermuth, N.: Graphical Models for Association Between Variables, Some of Which are Qualitative and Some Quantitative. Annals of Statistics 17, 31–57 (1989)

[14] Nielsen, T.D., Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer (2007)

[15] Peña, J.M.: Approximate Counting of Graphical Models Via MCMC. In: Proceedings of the 11th International Conference on Artificial Intelligence and Statistics. pp. 352–359 (2007)

[16] Perlman, M.: Graphical Model Search Via Essential Graphs. Tech. rep., University of Washington, Technical Report 367 (2000)

[17] Robinson, R.W.: Counting Labeled Acyclic Digraphs. New Directions in the Theory of Graphs pp. 239–273 (1973)

[18] Robinson, R.W.: Counting Unlabeled Acyclic Digraphs. In: Proceedings of the Fifth Australian Conference on Combinatorial Mathematics. pp. 28–43 (1977)
[19] Sonntag, D., Peña, J.M.: Learning Multivariate Regression Chain Graphs under Faithfulness. In: Proceedings of the 6th European Workshop on Probabilistic Graphical Models. pp. 299–306 (2012)
[20] Sonntag, D., Peña, J.M.: Chain Graph Interpretations and Their Relations. In: Proceedings of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty (ECSQARU 2013). pp. 510–521 (2013)
[21] Steinsky, B.: Enumeration of Labelled Chain Graphs and Labelled Essential Directed Acyclic Graphs. Discrete Mathematics pp. 266–277 (2003)
[22] Volf, M., Studený, M.: A Graphical Characterization of the Largest Chain Graphs. International Journal of Approximate Reasoning pp. 209–236 (1999)
[23] Wermuth, N., Sadeghi, K.: Sequences of Regression and Their Independences. Invited paper in TEST (2012)
[24] Wright, E.M.: A Relationship between Two Sequences. In: Proceedings of the London Mathematical Society. pp. 296–304 (1967)