

Aprendizado de máquina e causalidade no auxílio ao tratamento de casos suicidas em pacientes infanto-juvenis

Isis F. Carvalho

Orientadora: Gisele L. Pappa

Departamento de Ciência da Computação

Universidade Federal de Minas Gerais

Belo Horizonte, Brasil

isisfcarvalho@ufmg.br

Abstract—O comportamento suicida em crianças e adolescentes é uma questão crítica de saúde pública, sendo a principal causa de mortalidade entre jovens em países desenvolvidos. Este estudo visa desenvolver modelos de aprendizado de máquina causais para prever e intervir na automutilação e suicídio nessa faixa etária, superando limitações dos modelos preditivos existentes que focam em correlações em vez de causalidade. Inicialmente, o projeto envolve a construção de um grafo causal baseado em dados observacionais, que servirá como base para a inferência contrafactual, uma abordagem promissora para personalizar intervenções terapêuticas. Apesar dos avanços em modelos causais, a descoberta causal mostrou-se desafiadora devido à representação insuficiente dos dados. O grafo final proposto, baseado em conhecimento especializado e ajustado com dados reprocessados, integra variáveis-chave relacionadas ao comportamento suicida. Este trabalho destaca a necessidade de validação por especialistas e aponta para futuras etapas no desenvolvimento de modelos para calcular contrafactuais e aprimorar intervenções. O objetivo é proporcionar uma abordagem mais eficaz e personalizada na prevenção de comportamentos suicidas, potencialmente salvando vidas jovens em risco.

I. INTRODUÇÃO

O comportamento suicida em crianças e adolescentes representa uma das questões de saúde pública mais alarmantes da atualidade. A cada 40 segundos, uma pessoa morre por suicídio no mundo, impactando entre 60 e 135 pessoas ao seu redor com essa perda irreparável. Especificamente, o suicídio é a principal causa de mortalidade entre jovens em países desenvolvidos [1] e figura como a quarta principal causa de morte na faixa etária de 15 a 29 anos [2]. Entre os transtornos mentais que afetam essa faixa etária, destacam-se a depressão, a ansiedade e os transtornos comportamentais [1].

A investigação dos fatores que conduzem ao comportamento suicida é crucial para prevenir esse trágico desfecho. No entanto, prever tendências suicidas é uma tarefa desafiadora, mesmo para especialistas, devido à complexidade dos fatores sociais e ambientais que influenciam a saúde mental. Esse desafio é ainda mais acentuado em contextos de desigualdade social, como no Brasil.

Transtornos como automutilação e comportamento suicida estão se tornando mais prevalentes em hospitais pediátricos,

com variabilidade significativa na gravidade dos casos [3]. A população infanto-juvenil é frequentemente negligenciada nos estudos psiquiátricos, destacando a necessidade de identificar fatores de risco específicos para suicídio e automutilação nessa faixa etária. Esse reconhecimento é essencial para implementar sistemas de triagem e tratamento precoce, direcionando pacientes de risco para intervenções adequadas.

Dada a crescente prevalência de comportamentos suicidas entre jovens, torna-se imperativo desenvolver estratégias que permitam uma melhor compreensão das causas subjacentes a esses comportamentos e a intervenção eficaz para evitá-los. A prevenção de suicídio em crianças e adolescentes não só salva vidas, mas também reduz o impacto emocional e social significativo que essas tragédias impõem a famílias, comunidades e sistemas de saúde.

Apesar dos avanços em modelos preditivos baseados em aprendizado de máquina (AM), os progressos têm sido limitados devido a desafios [4] como a baixa prevalência de casos, o desbalanceamento de dados e a complexidade dos determinantes dos distúrbios mentais, que resultam de interações multifatoriais entre fatores genéticos e ambientais [5]. A maioria dos modelos atuais foca em correlações entre fatores e desfechos, sem abordar a causalidade, o que limita sua eficácia em orientar intervenções terapêuticas.

Este projeto tem como objetivo final propor novos modelos de aprendizado de máquina causais para a predição de automutilação e suicídio em crianças e adolescentes. Esses modelos permitirão o uso de técnicas de inferência contrafactual para orientar intervenções terapêuticas individualizadas. A inferência contrafactual possibilita avaliar os resultados de intervenções com base em dados observacionais, oferecendo vantagens significativas em relação aos estudos clínicos randomizados (ECRs), que são caros, eticamente questionáveis em psiquiatria e limitados em sua abrangência. O desenvolvimento desses modelos será fundamentado em um conjunto de dados composto por preditores clínicos e demográficos, visando aprimorar a precisão e a aplicabilidade das predições e intervenções terapêuticas.

Dessa forma, espera-se que os modelos desenvolvidos con-

tribuem para uma abordagem mais eficaz na prevenção de comportamentos suicidas, proporcionando intervenções mais precisas e personalizadas, e, conseqüentemente, salvando vidas jovens em risco.

O presente trabalho está organizado em 6 seções. Introduzimos o problema na Seção I. Na Seção II, temos a exposição do referencial teórico necessário para compreender o trabalho como um todo. A Seção III conta com a descrição dos dados que serão a base do projeto. A Seção IV traz a descrição do desenvolvimento do projeto e os resultados dos experimentos realizados. A Seção V aborda uma discussão dos resultados obtidos ao fim dos experimentos. Por último, a Seção VI conclui o trabalho com uma breve passada pelos próximos passos no projeto.

II. REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS

A Teoria da Causalidade de Judea Pearl [6, 7, 8, 9] é um marco no campo da inferência causal, oferecendo ferramentas e conceitos para modelar, identificar e inferir relações causais a partir de dados observacionais. Esta teoria é fundamental para o desenvolvimento de modelos de aprendizado de máquina capazes de prever e prevenir comportamentos suicidas em crianças e adolescentes, utilizando técnicas avançadas de inferência contrafactual para orientar intervenções terapêuticas personalizadas.

A. Escada da Causalidade

A Escada da Causalidade é uma hierarquia proposta por Pearl que classifica diferentes níveis de capacidade de raciocínio causal:

1) Associação

O nível mais básico, alcançado pela estatística tradicional, focado em detectar correlações e dependências estatísticas entre variáveis. Modelos nesse nível respondem a perguntas do tipo “*O que está acontecendo?*”, através de observação direta dos dados.

Exemplo: A identificação de uma correlação entre o uso excessivo de redes sociais e sintomas de depressão em adolescentes.

2) Intervenção

Este nível envolve a compreensão dos efeitos de manipulações diretas sobre as variáveis (que chamamos de intervenções). Modelos nesse nível respondem a perguntas do tipo “*O que acontece se eu fizer X?*” e requerem uma compreensão mais profunda das relações causais além da simples correlação.

Exemplo: Avaliar o impacto de programas de apoio psicológico na redução de comportamentos suicidas em adolescentes.

3) Contrafactuais

O nível mais avançado, que permite responder a perguntas hipotéticas como “*O que teria acontecido se eu tivesse feito Y em vez de X?*”. Este nível é crucial para a personalização de intervenções, simulando e avaliando os efeitos de diferentes ações sobre os

resultados desejados.

Exemplo: Analisar como diferentes abordagens terapêuticas poderiam ter alterado o comportamento de um adolescente que cometeu automutilação.

O objetivo desse projeto é obter um modelo causal que consegue responder perguntas contrafactuais: se entendemos o porquê de um certo paciente ter desenvolvido comportamento suicida, temos o poder de tomar medidas de tratamento mais adequadas no futuro, de forma a mitigar o sofrimento mental do paciente.

B. Structural Causal Models (SCMs)

Um modelo estrutural causal (ou SCM), proposto por [6], é uma ferramenta matemática para modelar mecanismos causais. Um SCM consiste de uma 4-tupla, cujos elementos são:

- Conjunto U de variáveis exógenas: Variáveis que dependem de fatores externos ao modelo.
- Conjunto V de variáveis endógenas: Variáveis determinadas por outras variáveis do modelo.
- Conjunto F de funções: Funções que descrevem as relações causais entre as variáveis, determinando como cada variável é influenciada por outras variáveis no sistema.
- $P(U)$: Função de probabilidade definida no domínio das variáveis exógenas, que captura a incerteza e a variabilidade não explicada pelas variáveis observadas.

Um SCM pode ser representado por um conjunto de equações estruturais, que explicitam as dependências causais entre as variáveis. Essas equações permitem a análise de intervenções e a inferência de contrafactuais. De acordo com [7], perguntas contrafactuais podem ser respondidas apenas por um SCM.

C. Grafos Causais

Um SCM pode também ser representado por um grafo direcionado acíclico (DAG), em que as variáveis em U e V são os nós do grafo, e as arestas que conectam os nós são os mecanismos causais F , que conectam as causas aos seus efeitos.

Os grafos causais são obtidos através de:

- **Conhecimento de Domínio:** Utilização de informações prévias e expertise sobre as relações causais entre variáveis.
- **Métodos de Inferência:** Algoritmos que aprendem a estrutura causal a partir dos dados observacionais, baseando-se em independências condicionais e outras técnicas estatísticas.

D. Construção de Modelos Causais Contrafactuais

A construção de modelos causais contrafactuais envolve os seguintes passos:

- **Definição do Modelo Causal:** Construir um grafo causal que represente as relações entre variáveis, utilizando conhecimento de domínio e/ou métodos de inferência.

- **Especificação das Funções Estruturais:** Determinar as equações que descrevem como cada variável é influenciada pelas outras variáveis no grafo causal.
- **Simulação de Intervenções:** Utilizar o SCM para simular intervenções específicas e calcular os efeitos diretos e indiretos dessas intervenções.
- **Cálculo de Contrafactuais:** Usar as equações estruturais para calcular cenários hipotéticos, avaliando os resultados de intervenções alternativas. Isso permite identificar as intervenções mais eficazes para prevenir comportamentos suicidas.

Um dos principais estudos relacionados ao objetivo deste trabalho é apresentado em [10], também no contexto da saúde mental. Este estudo utiliza técnicas de *deep learning* para aprender associações e modelar os relacionamentos causais entre as variáveis, resultando na construção de uma rede neural capaz de calcular contrafactuais.

III. CONJUNTO DE DADOS

O conjunto de dados utilizado neste projeto abrange 2.365 prontuários de pacientes admitidos no Centro Psíquico de Adolescência e de Infância (CEPAI), com alguns indivíduos apresentando múltiplas admissões, resultando em um total de 1.720 pacientes únicos [11]. A faixa etária dos pacientes no banco de dados de emergência psiquiátrica varia de 1 a 18 anos, e os dados foram coletados de junho de 2017 a maio de 2018. Este conjunto de dados inclui uma rica variedade de informações, compreendendo 27 características pessoais e sociodemográficas (tais como raça, gênero, local de nascimento, localização residencial, situação escolar, situação de moradia e responsável legal pelos pacientes) e 123 características clínicas, que abrangem detalhes sobre o motivo da busca por assistência psiquiátrica, histórico familiar de transtornos mentais, informações sobre abuso de substâncias, diagnósticos psiquiátricos recebidos após o atendimento na unidade, atrasos no desenvolvimento neuropsicomotor e eventos traumáticos anteriores, entre outros aspectos relevantes.

Foram excluídas 75 admissões em que os pacientes: (i) não tinham nenhuma informação (5); (ii) tinham apenas informações pessoais (37); (iii) não possuíam informações sobre as motivações para buscar ajuda ou diagnósticos (31); (iv) não completaram as etapas de triagem no centro (1); ou (v) não tinham informação de idade (1). Em relação às variáveis, 12 das 150 continham informações textuais (por exemplo, uma variável contendo o motivo escrito pelo qual o paciente saiu da escola). Em média, mais de 87% dos dados nessas variáveis estavam ausentes, e por isso elas foram descartadas. Outras 40 variáveis não possuíam conteúdo algum e também foram descartadas. O conjunto de dados original também continha 21 características relacionadas ao comportamento suicida, que apresentavam, em sua maioria, valores ausentes (mais de 80% das instâncias) e, portanto, foram excluídas devido ao alto risco de imputação de dados nesse contexto.

Das 77 características restantes, com a assistência de profissionais de psiquiatria, 57 foram consideradas mais informativas para as tarefas em questão. Examinamos as

seguintes variáveis como predictoras nos modelos de aprendizado de máquina: motivação para buscar ajuda psiquiátrica, diagnóstico recebido no CEPAL, local de nascimento, local de residência, identidade racial, situação escolar, tipo e tempo de serviço no CEPAL, histórico familiar de fatores psiquiátricos, diagnóstico previamente recebido, experiências traumáticas, atraso no desenvolvimento, número de pessoas no domicílio, idade, primeira vez no CEPAL, gênero e status de hospitalização.

Das 57 características selecionadas, 33 eram categóricas. Essas características categóricas, como por exemplo as cinco colunas contendo os diagnósticos recebidos por um paciente no CEPAL, foram binarizadas utilizando um dicionário de dados, resultando em 21 novas características binárias correspondentes aos 21 possíveis diagnósticos (por exemplo, se o paciente tinha depressão, esquizofrenia ou dificuldades de aprendizagem). O conjunto de dados final, após todos os tratamentos, consiste em 2.289 admissões para 1.687 pacientes únicos (1.071 do sexo masculino e 616 do sexo feminino), com cada admissão descrita por 154 características.

Esses dados serão a base para o desenvolvimento do projeto, a partir dos quais planejamos calcular os efeitos das possíveis intervenções e os tão desejados contrafactuais.

IV. DESENVOLVIMENTO DO PROJETO

Como definido na subseção II-D, o passo primordial para a construção de um modelo causal contrafactual é a definição do grafo que representa as relações causais entre as variáveis no contexto do problema, tendo em mente os dados disponíveis. Este passo é fundamental, pois configura a base para todo o trabalho subsequente, sendo crucial que o grafo represente corretamente o conhecimento existente sobre as relações causais das variáveis.

Entretanto, determinar a conexão causal entre as variáveis a partir de dados observacionais é uma tarefa desafiadora, visto que essas relações nem sempre são evidentes nos dados disponíveis. Além disso, podem existir caminhos causais não triviais entre as variáveis, ou os dados podem não conter informações essenciais para as relações causais propostas.

Seguindo o planejamento delineado na proposta do projeto, as atividades conduzidas durante o período da POC I focaram exclusivamente na obtenção de tal grafo. A presente seção descreverá, em detalhes, os métodos e caminhos explorados para alcançar este objetivo.

A. Algoritmos para Descoberta Causal

Aprender a estrutura causal, representada por um grafo acíclico direcionado (DAG), a partir de dados observacionais define o que chamamos de Descoberta Causal. Este problema é considerado NP-difícil [12], devido à natureza combinatória do espaço de busca dos DAGs.

Os algoritmos de descoberta causal se baseiam em um conjunto de dados e calculam relações estatísticas entre variáveis, estimando a força causal dessas relações. Esses algoritmos podem ser classificados em duas abordagens principais [13]:

- **Constraint-based:** Esta abordagem assume que o grafo desejado deve refletir todas as independências condicionais presentes nos dados. Testes estatísticos de independência condicional são realizados para determinar quais grafos são consistentes com os dados. Embora sejam precisos, esses algoritmos são computacionalmente intensivos e requerem grandes volumes de dados.
- **Score-based:** Esta abordagem busca encontrar um grafo que maximize uma função *score*. Como o grafo deve ser um DAG para modelar as relações causais corretamente, o problema de otimização possui uma restrição que garante a aciclicidade do grafo. Esses algoritmos são mais simples e eficazes, mas enfrentam desafios devido ao espaço de busca superexponencial dos DAGs.

Neste projeto, consideramos apenas algoritmos baseados em *score*, pois são atualmente considerados o estado da arte na área de descoberta causal. Os algoritmos estudados incluem NOTEARS [14], DAGMA [15], e DAGMA-DCE [13].

O algoritmo NOTEARS [14], proposto em 2018, reformula o problema de descoberta causal como um problema de otimização contínua, evitando as complicações relacionadas ao espaço de busca combinatorial dos DAGs. Ele consegue isso ao propor uma nova caracterização de aciclicidade do grafo, mais eficiente do que as abordagens anteriores.

O algoritmo DAGMA¹ [15], proposto em 2022, baseia-se na mesma abordagem do NOTEARS, mas inova ao propor uma nova função de aciclicidade que oferece várias vantagens em relação ao algoritmo anterior. A mais significativa é que a nova definição de aciclicidade pode ser computada aproximadamente uma ordem de grandeza mais rápido do que a função antiga.

Por fim, o algoritmo DAGMA-DCE² [13], proposto em 2024, se baseia no DAGMA, como o próprio nome sugere, mas introduz uma medida de força causal interpretável para definir o grafo causal que otimiza a função *score*.

Durante os estudos iniciais do projeto, o algoritmo NOTEARS foi descartado por não ser invariante à escala, o que o torna inadequado para identificar relações causais a partir dos dados reais [16]. Portanto, seguimos com os algoritmos DAGMA e DAGMA-DCE.

B. Experimentos com DAGMA e DAGMA-DCE

Para obter grafos gerados pelos algoritmos DAGMA e DAGMA-DCE, foi necessário inicialmente implementar scripts que recebessem os dados observacionais e retornassem o grafo causal obtido pelo algoritmo na forma de uma matriz de adjacências. Em seguida, desenvolvemos um script adicional para gerar visualizações dos grafos a partir dessas matrizes, permitindo a análise e comparação dos resultados experimentais.

Com esses scripts em mãos, realizamos diversos experimentos de geração de grafos causais, variando o recorte de dados utilizado. Foram considerados três recortes de features,

aplicados a ambos os algoritmos: todas as features disponíveis (denominado *all_feats*), todas as features binárias disponíveis (*all_binary_feats*) e features consideradas mais importantes por especialistas da área (como motivações para procurar atendimento, diagnósticos recebidos, sexo, idade, etc.), denominado *subset*. A seguir, discutimos os resultados obtidos para cada algoritmo.

O algoritmo DAGMA³ permite a obtenção de grafos causais baseados em Modelos Causais Estruturais (SCMs) lineares e não lineares, com suporte para dados contínuos em ambos os casos. Para dados binários, o suporte é limitado ao caso linear, onde é possível utilizar a função logística como *score*. Considerando que nossos dados são majoritariamente binários, testamos ambos os casos (linear e não-linear) para observar os resultados, mas focamos principalmente no caso linear, que se ajusta melhor aos dados disponíveis.

Para o caso linear, inicializamos um objeto da classe `DagmaLinear`, especificando o *score* utilizado e obtendo o grafo com a função `fit`. É possível também especificar, nesta função, arestas que devem ser incluídas ou excluídas no grafo final. No entanto, a implementação atual da biblioteca não garante que todas as arestas listadas para inclusão estarão no grafo final.

Para o caso não linear, primeiramente inicializamos um modelo de equações representado por uma rede neural multicamadas (MLP) com a classe `DagmaMLP`, especificando o número de nós por camada. Em seguida, inicializamos um objeto da classe `DagmaNonlinear`, passando esse modelo de equações como parâmetro, e utilizamos a função `fit` desse objeto para obter o grafo final.

A Figura 1 apresenta um trecho de um grafo obtido utilizando a classe `DagmaLinear` com o recorte de dados *subset*. Observa-se que algumas arestas conectam elementos chave relacionados ao comportamento suicida de maneira contraintuitiva: por exemplo, “TAE” (tentativa de autoextermínio) é mostrado como causa de “Sintomas de Depressão”, que por sua vez é causa de “Idade”. Além disso, “Automutilação” é representada como causa de “Diagnóstico de Depressão”, que novamente leva a “Idade”. A Figura 1 é ilustrativa desses resultados inconsistentes retornados pelo `DagmaLinear`.

De maneira semelhante, a Figura 2 mostra um trecho do grafo obtido com a classe `DagmaNonlinear` e o recorte de dados *subset*. Novamente, observa-se relações estranhas, como “TAE” e “Ideação de Autoextermínio” levando a “Sintomas de Depressão”. No entanto, é possível observar à direita um encadeamento mais coerente com a psicologia pediátrica: “Agitação” leva a “Nervosismo/irritabilidade”, que leva a “Agressividade”.

Esses resultados ilustram as complexidades e desafios envolvidos na descoberta causal a partir de dados observacionais, ressaltando a importância de validação e ajuste contínuo dos modelos com base no conhecimento especializado da área.

Diante dos resultados um tanto quanto decepcionantes, voltamos nossa atenção ao algoritmo DAGMA-DCE.

¹Repositório do DAGMA no GitHub.

²Repositório do DAGMA-DCE no GitHub.

³Documentação detalhada do DAGMA.

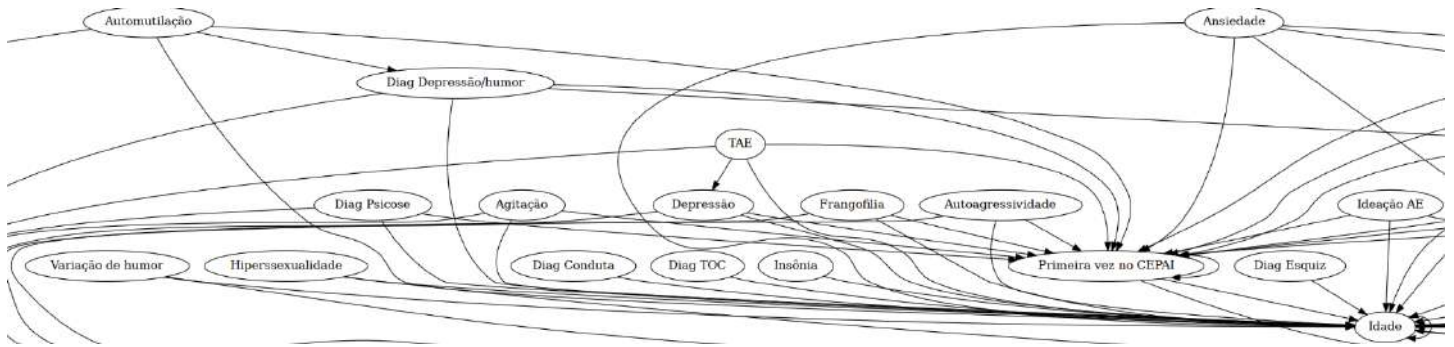


Fig. 1. Trecho de grafo obtido com DagmaLinear, com recorte de dados *subset*.

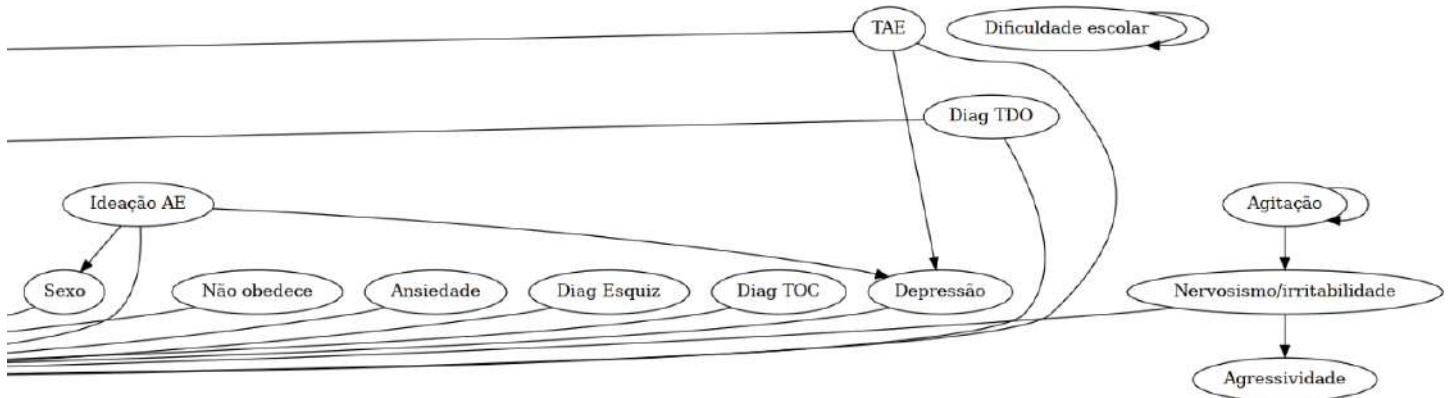


Fig. 2. Trecho de grafo obtido com DagmaNonlinear, com recorte de dados *subset*.

O DAGMA-DCE é implementado com base no DAGMA, funcionando de maneira semelhante à descrita anteriormente. No entanto, uma particularidade deste algoritmo é a ausência de uma versão linear, existindo apenas a versão não linear. Considerando que um dos experimentos no repositório do DAGMA-DCE utiliza sua classe não linear para testar um problema de SCMs lineares, com resultados superiores ao DAGMA, decidimos também testar o DAGMA-DCE não linear.

Um diferencial interessante do DAGMA-DCE é que ele utiliza o DAGMA para inicializar os pesos do modelo de equações, realizando suas próprias atualizações subsequentes com o método *fit*.

A Figura 3 ilustra um trecho do grafo gerado utilizando a classe *DagmaDCE* com o recorte de dados *subset*. Observa-se que todas as relações têm como efeito a variável "Idade", o que não faz sentido e nem é útil em nosso contexto, visto que é mais provável que a idade seja a causa de alguns fatores, e não o contrário. Além disso, a variável idade não pode sofrer intervenções, uma vez que não podemos alterar a idade de um indivíduo no presente. Portanto, embora o grafo gerado seja mais limpo, ele não nos fornece informações relevantes de causalidade para serem usadas nas próximas etapas do projeto. O mesmo pode ser observado na Figura 4, que apresenta apenas relações causais entre variáveis altamente correlacionadas e pouco informativas.

Diante dos problemas observados nos resultados obtidos até então, hipotetizamos uma série de possibilidades sobre os motivos pelos quais os algoritmos podem não estar respondendo bem aos dados, bem como alternativas para sanar esses problemas. Uma de nossas principais suposições, explorada em trabalhos anteriores [17], é que os dados são pouco representativos dos mecanismos causais que desejamos captar. Como apresentado na Seção III, temos pouco mais de 2 mil instâncias, com muitas colunas trazendo informações semelhantes. É possível também que, durante o processamento dos dados com one-hot encoding, algumas correlações espúrias tenham sido introduzidas, o que pode dificultar para os algoritmos a identificação de relações causais com base em uma função de *score* puramente estatística.

Embora seja impossível gerar mais dados ou torná-los mais representativos, podemos mitigar alguns problemas inerentes ao conjunto de dados. Nesse sentido, o tratamento dos dados foi revisado e algumas decisões foram tomadas para obter uma base mais concisa e informativa. Por exemplo, removemos diversas colunas que não traziam nenhuma informação útil ao problema em questão (como o tipo e o turno em que o paciente recebeu o atendimento) e consolidamos as colunas de diagnósticos prévios com os diagnósticos recebidos durante o atendimento, unindo essa informação em variáveis mais concisas. Removemos também a variação de algumas colunas, que antes eram bastante detalhadas, como a de diagnóstico

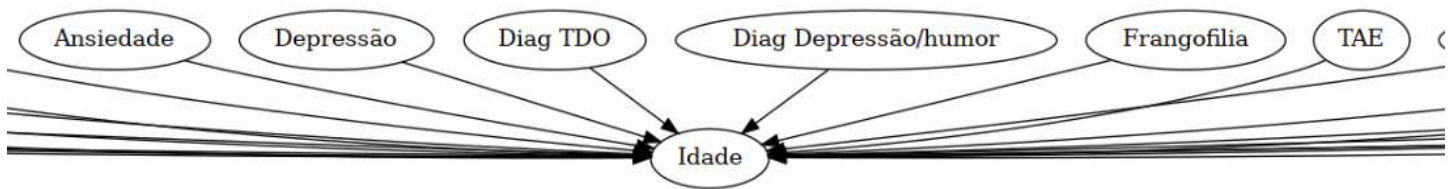


Fig. 3. Trecho de grafo obtido com DagmaDCE, com recorte de dados *subset*.



Fig. 4. Trecho de grafo obtido com DagmaDCE, com recorte de dados *allfeats*.

clínico, mantendo apenas uma variável binária que indica se o paciente tem registro de qualquer quadro clínico problemático. Após esse processo, reduzimos a base para 62 colunas. Essa redução também melhora o tempo de resposta dos algoritmos de descoberta causal, que passam a lidar com menos combinações de arestas possíveis. Para posterior identificação, esse recorte de dados será chamado de *reprocess*.

Após esse reprocessamento dos dados, dois outros caminhos para melhorar os resultados se apresentaram. O primeiro foi a introdução de conhecimento de domínio: poderíamos iniciar a descoberta causal a partir de um grafo que faça sentido para a área da psico-pediatria e que retrate algumas relações causais já estabelecidas. Discutiremos esse desenvolvimento na Seção IV-C. O segundo caminho possível seria a implementação de uma função de *score* personalizada para obtenção dos grafos causais, em resposta à inabilidade de ambos os algoritmos em lidar simultaneamente com dados mistos (contínuos e categóricos). Esse *score* poderia ser tanto o Bayesian Information Criterion (BIC) como o Minimum Description Length (MDL), adaptados para dados categóricos pela utilização da medida Bayesian Dirichlet equivalent uniform (BDeu). No entanto, todas as variáveis resultantes do novo tratamento de dados são binárias. Por conta disso, essa alternativa, também devido à sua grande complexidade de integração com a estrutura da biblioteca DAGMA já existente, foi despriorizada por ora, sendo uma opção para a próxima etapa do projeto, caso necessário.

C. Inclusão de Conhecimento de Domínio

Como mencionado na subseção anterior, uma alternativa robusta para mitigar o problema dos algoritmos de descoberta causal não abordarem as relações mais importantes é simplesmente forçar a inclusão de algumas arestas desde o início. Contudo, infelizmente esse recurso está disponível apenas para a classe `DagmaLinear`, por meio do parâmetro `include_edges` da função `fit`. Como já dito antes, a implementação atual não garante que as arestas especificadas para inclusão estarão realmente presentes no grafo final. Portanto, os experimentos a partir deste ponto serão um pouco

mais limitados, mas espera-se que os resultados sejam mais alinhados com as expectativas.

O grafo que utilizaremos como base para a descoberta causal daqui em diante foi proposto por profissionais da área da psico-pediatria e estruturado de maneira a corresponder às variáveis disponíveis no conjunto de dados. A Figura 5 apresenta esse grafo por completo. Observa-se que as conexões entre as variáveis fazem mais sentido do que qualquer outro resultado observado até agora. Esse grafo, por si só, já seria suficiente para dar continuidade ao trabalho, visto que ele contém o encadeamento causal das variáveis conforme os profissionais observam no tratamento diário de pacientes. Entretanto, é possível que os algoritmos de descoberta causal possam agregar conhecimento não trivial ao grafo em questão, o que justifica o aprofundamento dos experimentos.

O resultado do primeiro experimento com a tentativa de inclusão das arestas representadas na Figura 5 pode ser observado na Figura 6. Com os parâmetros originais e utilizando o recorte *reprocess*, o algoritmo falhou tanto em manter as relações do grafo inicial quanto em fornecer novas relações que sejam úteis no contexto de descoberta causal. Observa-se muitas relações entre variáveis relativas a dados sociodemográficos dos pacientes, mas poucas relações causais capturadas entre os fatores clínicos e psiquiátricos que têm maior evidência para o aparecimento de comportamento suicida.

Diante desse resultado, foi feito um novo recorte de dados, removendo qualquer feature que diga respeito à cor, local de nascimento ou residência do paciente, além de variáveis relativas às características do atendimento recebido no CEPAI. Esse novo recorte será identificado como *best-cut*. O objetivo é focar nas variáveis mais relevantes para o comportamento suicida, eliminando possíveis fontes de ruído que possam ter prejudicado os resultados anteriores.

Na Figura 7, é possível observar uma mudança no comportamento dos resultados. Embora as variáveis-chave para comportamento suicida, como “TAE” e “Ideação AE”, não estejam presentes, variáveis como “Ansiedade” e “Sexo” aparecem conforme o esperado. Na Figura 8, o threshold de inclusão de arestas, um parâmetro da função `fit`, foi ajustado de 0.3

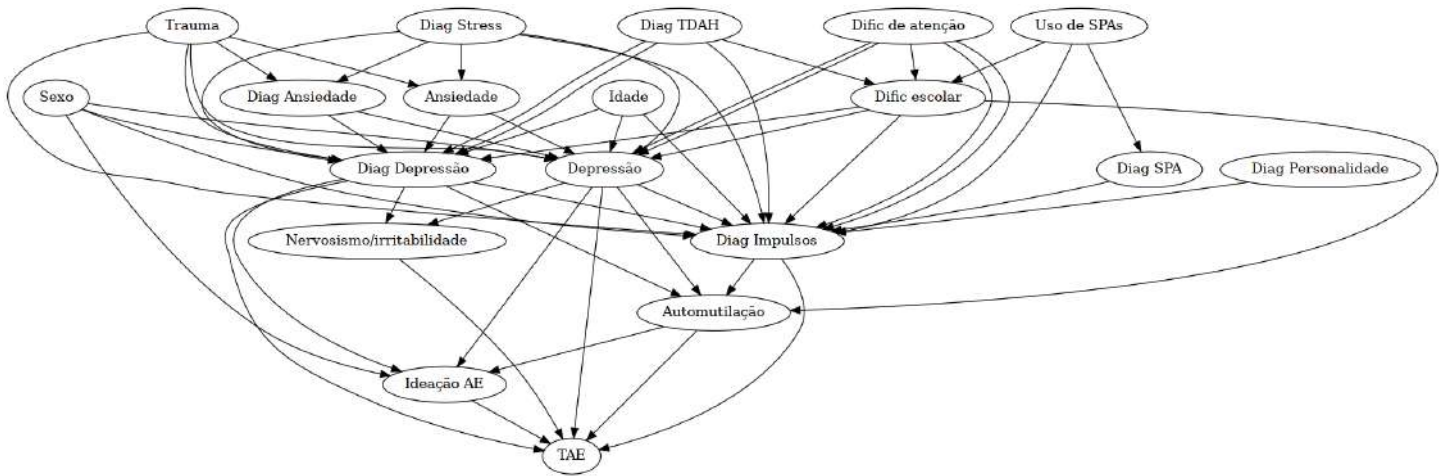


Fig. 5. Grafo mapeando as variáveis relacionadas a comportamentos suicidas na infância e adolescência e suas relações causais, proposto por profissionais da área da psico-pediatria.

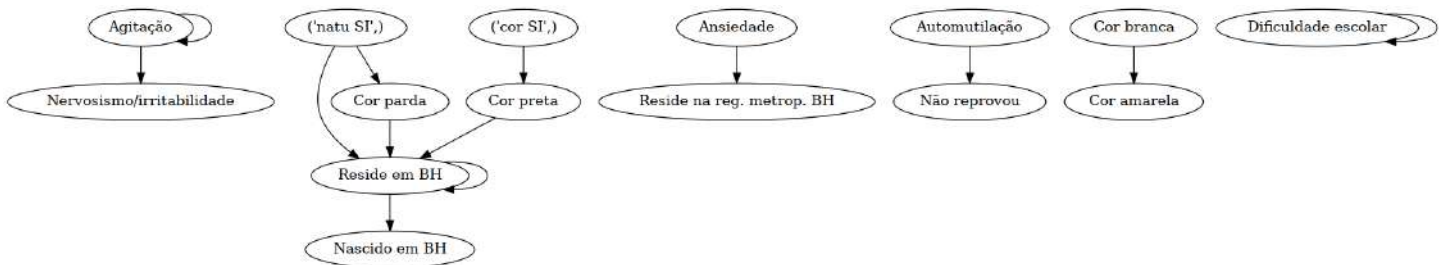


Fig. 6. Grafo obtido com DagmaLinear, com inclusão de arestas de conhecimento de domínio, utilizando o novo recorte de dados *reprocess*.

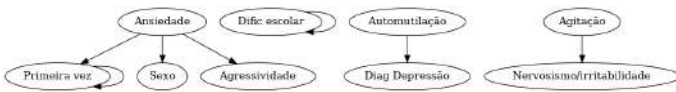


Fig. 7. Grafo obtido com DagmaLinear, com inclusão de arestas de conhecimento de domínio, utilizando o novo recorte de dados *best-cut*.

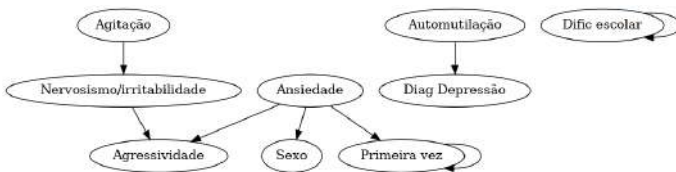


Fig. 8. Grafo obtido com DagmaLinear, com inclusão de arestas de conhecimento de domínio e diminuição de threshold de inclusão de arestas, utilizando o novo recorte de dados *best-cut*.

para 0.1, a fim de verificar se conseguimos obter arestas mais relevantes para o problema.

Observa-se que praticamente não houve mudanças, o que sugere que os problemas são devidos, em parte, aos dados pouco representativos e às limitações da classe DagmaLinear. Com isso em mente, realizamos experimentos com as classes DagmaNonlinear e DagmaDCE para avaliar se, mesmo sem iniciar a partir de um grafo baseado em conhecimento de domínio, seria possível captar

relações causais importantes entre as variáveis disponíveis. No entanto, os experimentos com a classe DagmaDCE não foram bem-sucedidos, pois, para todos os thresholds testados, o grafo retornado estava vazio. Portanto, os experimentos apresentados nas Figuras 9 e 10 foram obtidos com a classe DagmaNonlinear, variando o threshold de inclusão de arestas e utilizando o último recorte de dados, denominado *best-cut*.

Na Figura 9, são visíveis várias relações causais que são plausíveis do ponto de vista da psico-pediatria. Para facilitar a análise, as arestas que são consistentes com o conhecimento da área estão marcadas em verde, enquanto aquelas que deveriam estar invertidas estão marcadas em vermelho. Embora o grafo apresente muitas arestas importantes, observa-se que as variáveis relacionadas ao comportamento suicida estão relativamente isoladas no lado esquerdo do grafo. Esse isolamento não é ideal para o objetivo deste trabalho, que é calcular os efeitos de intervenções visando a prevenção do comportamento suicida.

Finalmente, a Figura 10 mostra um subconjunto das arestas com maior força causal identificadas na Figura 9. Embora várias das relações apresentadas sejam relevantes, o isolamento das variáveis “Automutilação”, “Ideação AE” e “TAE” persiste, semelhante ao que foi observado no grafo anterior.

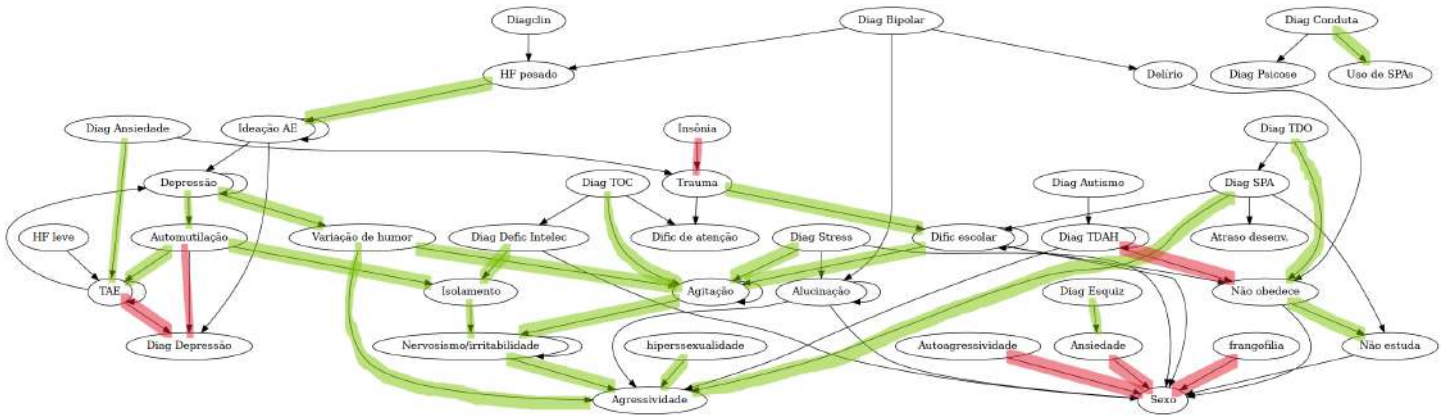


Fig. 9. Grafo obtido com DagmaNonlinear, com threshold de inclusão de arestas 0.2, utilizando o recorte de dados *best-cut*.



Fig. 10. Grafo obtido com DagmaNonlinear, com threshold de inclusão de arestas 0.3, utilizando o recorte de dados *best-cut*.

V. DISCUSSÃO DOS RESULTADOS

O desenvolvimento deste trabalho seguiu um processo altamente iterativo, como detalhado na seção anterior. Esse processo envolveu uma fase experimental intensiva, caracterizada pela execução repetitiva de experimentos, análise dos resultados obtidos e formulação de hipóteses para aprimorar as iterações subsequentes. Durante essas etapas, foram implementados novos tratamentos de dados, ajustadas parametrizações de algoritmos e incorporado conhecimento de domínio.

No entanto, os resultados obtidos até o momento indicam que os algoritmos de descoberta causal utilizados não atingiram o desempenho esperado. A hipótese mais robusta para explicar essa performance insatisfatória é que os dados disponíveis não são suficientemente representativos para gerar um grafo causal que reflita com precisão os mecanismos causais subjacentes às relações entre as variáveis.

Dado esse problema, é inviável realizar qualquer manipulação adicional dos dados que possa melhorar significativamente o desempenho dos algoritmos de descoberta causal além dos resultados obtidos nos experimentos mais recentes.

Apesar dessas limitações, o objetivo final desta etapa do projeto é produzir um grafo causal que capture de maneira eficaz os mecanismos causais relacionados ao comportamento suicida em crianças e adolescentes. Considerando as restrições discutidas, o grafo apresentado na Figura 5, complementado com algumas relações identificadas nos experimentos que geraram a Figura 9, representa o melhor resultado possível de ser alcançado. Portanto, o grafo final proposto neste trabalho é exibido na Figura 11. Este grafo integra os pares de variáveis

“Depressão” e “Diag Depressão” e “Ansiedade” e “Diag Ansiedade”, além de incluir 10 arestas que foram apresentadas propostas na Figura 9, que são:

- HF pesado⁴ → Depressão
- Depressão → Variação de humor
- Automutilação → Isolamento
- Isolamento → Nervosismo/irritabilidade
- Diag Defic Intelec → Isolamento
- Diag TOC → Agitação
- Agitação → Nervosismo/irritabilidade
- Diag Stress → Agitação
- Dific escolar → Agitação
- Diag Conduta → Uso de SPAs

As relações são bastante autoexplicativas e agregam conhecimento causal ao grafo gerado por conhecimento de domínio.

VI. CONCLUSÕES

A inferência contrafactual oferece uma abordagem promissora para a personalização de intervenções terapêuticas, permitindo a avaliação de cenários hipotéticos com base em dados observacionais. Essa abordagem pode proporcionar insights valiosos sobre os efeitos de diferentes intervenções, superando as limitações dos estudos clínicos randomizados.

Para implementar a inferência contrafactual, o primeiro passo é obter um grafo que represente os mecanismos causais envolvidos no comportamento suicida em crianças e adolescentes. Este grafo servirá de base para o desenvolvimento

⁴HF pesado se refere a todos os registros históricos de transtornos e ocorrências de eventos trágicos na família do paciente. Esta feature foi caracterizada dessa forma para facilidade de representação, mas inclui, por exemplo, histórico de familiares usuários de substâncias psicoativas (SPAs), suicídio ou depressão na família, entre outros.

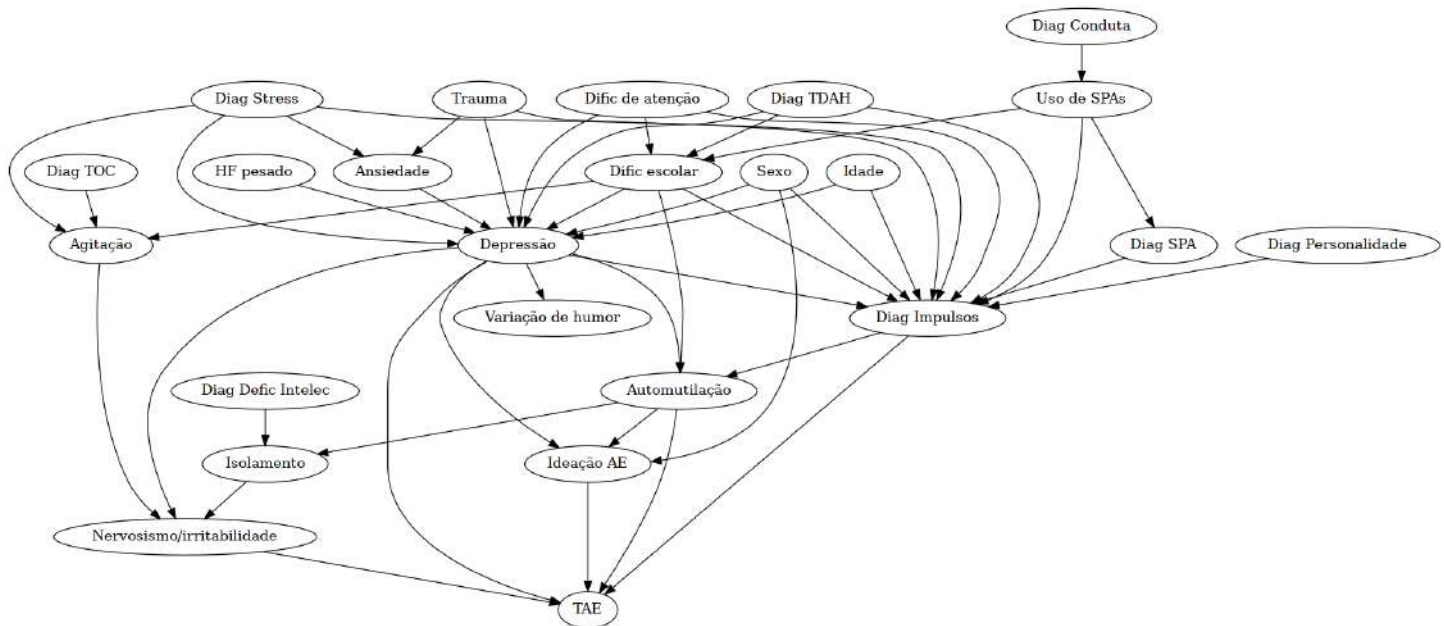


Fig. 11. Grafo final obtido frente ao conhecimento de domínio e aos resultados dos experimentos realizados.

do modelo causal necessário para calcular os efeitos de intervenções e, posteriormente, realizar a inferência contrafactual.

Para assegurar a qualidade do grafo, é essencial submetê-lo à revisão de especialistas na área. A aprovação desses especialistas, que possuem o conhecimento de domínio necessário, é crucial para validar as relações de causa e efeito propostas.

Com o grafo validado, o próximo passo, a ser abordado em futuros trabalhos, será a escolha e implementação de um modelo que permita alcançar os níveis 2 e 3 da escada da causalidade de Pearl. Isso possibilitará o cálculo de contrafactuais e a realização de análises mais detalhadas sobre o impacto das intervenções propostas.

REFERENCES

- [1] D. Knipe, P. Padmanathan, G. Newton-Howes, L. F. Chan, and N. Kapur, "Suicide and self-harm," *Lancet*, vol. 399, no. 10338, pp. 1903–1916, May 2022.
- [2] World Health Organization, "Suicide," <https://www.who.int/news-room/fact-sheets/detail/suicide>, 2021, accessed: 2024-04-02.
- [3] E. Maguire, C. McGrath, and P. Byrne, "P409 children in distress - a 10 year descriptive study of children aged 0-12 years referred for emergency psychiatric assessment in tallaght university hospital," *Archives of Disease in Childhood*, vol. 104, no. Suppl 3, pp. A319–A320, 2019.
- [4] C. M. McHugh and M. M. Large, "Can machine-learning methods really help predict suicide?" *Current opinion in psychiatry* vol. 33, 2020.
- [5] R. Uher and A. Zwicker, "Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness," *World psychiatry : official journal of the World Psychiatric Association (WPA)* vol. 16, 2017.
- [6] J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.
- [7] —, *Causality*, 2nd ed. Cambridge, UK: Cambridge University Press, 2009.
- [8] J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer*. John Wiley Sons, Incorporated, 2016.
- [9] J. Pearl and D. Mackenzie, *The Book of Why*. New York: Basic Books, 2018.
- [10] G. F. Marchezini, A. M. Lacerda, G. L. Pappa, W. Meira, D. Miranda, M. A. Romano-Silva, D. S. Costa, and L. M. Diniz, "Counterfactual inference with latent variable and its application in mental health care," *Data Mining and Knowledge Discovery*, vol. 36, pp. 811–840, 2022.
- [11] M. Lobato Machado, M. Hibner, D. Nogueira, M. Rezende, I. Bórem, L. da Cunha, and D. Marques de Miranda, "Irritability in an open-door pediatric psychiatric emergency service in a middle-income country," *Neuropsychiatrie de l'Enfance et de l'Adolescence*, vol. 70, no. 6, pp. 336–342, 2022.
- [12] D. M. Chickering, D. Heckerman, and C. Meek, "Large-sample learning of bayesian networks is np-hard," *J. Mach. Learn. Res.*, vol. 5, p. 1287–1330, dec 2004.
- [13] D. Waxman, K. Butler, and P. M. Djurić, "Dagma-dce: Interpretable, non-parametric differentiable causal discovery," *IEEE Open Journal of Signal Processing*, vol. 5, p. 393–401, 2024. [Online]. Available: <http://dx.doi.org/10.1109/OJSP.2024.3351593>
- [14] X. Zheng, B. Aragam, P. Ravikumar, and E. P. Xing, "Dags with no tears: Continuous optimization for structure learning," 2018. [Online]. Available: <https://arxiv.org/abs/1803.01422>

- [15] K. Bello, B. Aragam, and P. Ravikumar, “Dagma: Learning dags via m-matrices and a log-determinant acyclicity characterization,” 2023.
- [16] M. Kaiser and M. Sipos, “Unsuitability of notears for causal graph discovery,” 2021. [Online]. Available: <https://arxiv.org/abs/2104.05441>
- [17] I. Carvalho, D. Miranda, A. da Silva, A. Lacerda, W. M. Jr., M. Romano-Silva, M. Lobato, and G. Pappa, “Prediction of suicidal behaviors in hospitalized children and adolescents in middle-income countries: a case study of brazil,” in *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional*. Porto Alegre, RS, Brasil: SBC, 2023, pp. 1225–1236. [Online]. Available: <https://sol.sbc.org.br/index.php/eniac/article/view/25776>