

Geração automática de modelo de relações de pré-requisitos a partir de avaliação de larga escala brasileiras: um estudo preliminar

Bruno Elias Penteadó

¹Instituto de Ciências Matemáticas e da Computação – Universidade de São Paulo (USP)

Avenida Trabalhador São-carlense, 400 – Centro – São Carlos – SP – Brasil

brunopetneado@usp.br

***Abstract.** A domain model is a crucial component for intelligent learning environments, because it drives the diagnostics of what is being assessed e how it will be worked by the students. One of the essential components for this model is the proper ordering of the learning topics, usually made by domain experts. This initial work aims to use public data of the applications of diagnostic large-scale assessments to automatically create a cognitive model of the domain that can be used in intelligent learning environments. The data used is from SAEB 2013, both its assessment items and the students' responses, besides a state-of-the-art algorithm (POKS). The partial results point to a viable method to be explored, in the construction of intelligent learning environments.*

***Resumo.** Um modelo de domínio é um componente crucial para ambientes de aprendizagem inteligentes, pois norteia o diagnóstico do que está sendo avaliado e como ele será trabalhado pelos alunos. Um dos componentes essenciais para este modelo é a ordenação apropriada de seus tópicos de estudo, que são usualmente feitas por especialistas do domínio. Este trabalho inicial tem por objetivo usar os dados públicos das aplicações das avaliações diagnósticas de larga escala para gerar automaticamente um modelo cognitivo de domínio que possa ser usado em ambientes inteligentes de aprendizagem. Foram usados dados do SAEB 2013, tanto de seus itens avaliativos quanto das respostas dos alunos, além de um algoritmo estado da arte (POKS). Os resultados parciais apontam para algo factível a ser explorado, para a construção de ambientes inteligentes de aprendizagem.*

1. Introdução

Um bom plano de ensino tem por fundamento se basear em uma estrutura de conhecimento que possa fazer com que o aluno seja introduzido ao assunto e, gradativamente, vá adquirindo conhecimentos em tópicos mais complexos, conforme se aprofunda no currículo. Para ambientes de ensino inteligentes, como os sistemas tutores inteligentes ou os sistemas hipermídia adaptativos educacionais, este sequenciamento é

fundamental, pois eles se baseiam nesta ordenação para adaptarem suas ações conforme o aluno avança pelo seu conteúdo.

O planejamento dessas sequências de aprendizagem é chamado de *modelagem de domínio*. Ele parte de princípios de teorias instrucionais e pressupõe que os alunos geralmente aprendem habilidades em sequência, partindo de habilidades mais simples e avançando gradualmente para as mais complexas, reforçando o que já foi aprendido conforme novos conhecimentos são apresentados [Bruner, 1966; Reigeluth e Stein, 1983]. Além disso, este modelo de domínio serve de base para a criação do chamado *modelo de aluno*, uma representação do conhecimento do aluno sobre o currículo a ser ensinado e que é usado para a adaptação automática das estratégias de ensino. O modelo do aluno é comumente gerado como uma cópia do modelo de domínio, sendo atualizado pela avaliação da aquisição de conhecimento realizada pelo sistema e, com isso, influenciando como será feita a intervenção automática para o aluno.

Geralmente, esta estrutura dos conhecimentos, os tópicos de estudo e as relações de pré-requisitos entre eles, é definida por especialistas no assunto, que o fazem de modo manual, especificando os tópicos, as habilidades envolvidas e qual o sequenciamento adequado - por meio de seu julgamento ou seguindo alguma teoria instrucional [Vuong, Nixon e Towle, 2011]. Trata-se de um processo caro e demorado, pois envolve o trabalho de muitos especialistas para analisar o assunto e da concordância com um modelo único e, ainda assim, gerar modelos de pouca confiabilidade [Nathan e Koedinger, 2000]. Esse procedimento de estruturação é complexo, pois as habilidades trabalhadas são *traços latentes*, ou seja, não podem ser medidos de modo completamente observável. Outro ponto discutido na literatura é a falta de validações empíricas para atestar a aderência do modelo criado pelo especialista com os resultados obtidos pelos alunos em situações de sua aplicação [Vuong, Nixon e Towle, 2011].

Este trabalho tem como motivação auxiliar o processo de especificação da estrutura de habilidades, ao tomar como ponto de partida as habilidades já estabelecidas pelas avaliações de larga escala – as chamadas *matrizes de referência* – e aplicar um método que possa extrair automaticamente relações de pré-requisitos entre as habilidades avaliadas. Até onde sabemos, este processo não foi aplicado a resultados de avaliações de larga escala, que será explorado neste trabalho.

Ao tirar proveito da grande quantidade de dados de desempenho de alunos nas avaliações em larga escala nacionais, o objetivo é analisar os dados de respostas dos alunos às provas nacionais de larga escala, e aplicar um método que possa identificar uma possível estrutura de relações de pré-requisito subjacente entre as habilidades que compõem a matriz de referência dessas avaliações. Com isso, espera-se obter como resultado deste método um modelo cognitivo das habilidades e as respectivas relações de pré-requisito, detectadas automaticamente e que possam ser reutilizadas na concepção de sistemas inteligentes voltados para auxiliar o estudo para essas avaliações.

2. Contextualização

Uma das questões mais importantes no planejamento instrucional de um curso é o sequenciamento das atividades de aprendizagem disponíveis para o curso – a ordem e a organização das atividades de aprendizagem afetam como a informação é processada e

retida [van Patten, Chao e Reigeluth, 1986]. Como vimos, esta estruturação é um processo complexo e que raramente é validado empiricamente.

As avaliações nacionais de larga escala, como a Prova Brasil e o ENEM, apresentam um conjunto de habilidades que avaliam o que se espera que os alunos aprendam em um ciclo de ensino, compiladas nas chamadas *matrizes de referência*. No entanto, até onde sabemos, não existe na literatura algum estudo que procurou estabelecer relações de pré-requisito entre estas habilidades. Uma vez que isso seja evidenciado, seus resultados podem influenciar como as habilidades são trabalhadas durante os ciclos de ensino.

Essas avaliações usam frequentemente itens de múltipla escolha e são usadas para medir traços amplos, como proficiência em Matemática. Seus resultados são usados para finalidades como: subsidiar a formulação de políticas públicas, acompanhar a evolução da aprendizagem ao longo do tempo além de produzir indicadores de unidades escolares ou sistemas de ensino. No Brasil, a Prova Brasil e o SAEB são avaliações aplicadas ao final dos ciclos do ensino básico (5º e 9º anos do fundamental e 3º do médio), e são elaboradas a partir das matrizes de referência.

Para as avaliações de larga escala, a abordagem atual considera todos os conhecimentos dentro de apenas uma escala unidimensional que considera a dificuldade do item em sua ordenação. No entanto, dada a amplitude das habilidades, especialistas concordam haver outros tipos de relações entre as habilidades, que não essa simples ordenação de dificuldade [Clements e Sarama, 2004]. O processo de criação manual dessas dependências é altamente custoso e de grande discordância entre os especialistas que modelam o domínio.

Como uma área que pode contribuir para a descoberta dessas relações, a Mineração de Dados Educacionais (MDE) é uma área de pesquisa relativamente recente que combina diversas disciplinas para buscar compreender como os alunos aprendem e, assim, criar ferramentas que suportem esta aprendizagem.

Koedinger et al. [2015], apresenta uma taxonomia das aplicações da área de MDE e propõem as seguintes categorias: i) avaliação da cognição, aprendizagem e desempenho; ii) transferência da aprendizagem e descoberta de modelos cognitivos; iii) análise de estados afetivos, motivação e cognição e iv) a análise da linguagem e do discurso. Este trabalho se encaixa na segunda categoria, ao buscar encontrar estruturas nos dados disponíveis sem uma ideia a priori do que deve ser encontrado, buscando determinar estruturas que emergem naturalmente a partir dos dados.

Ao contrário dos diagnósticos amplos das avaliações de larga escala, a literatura de MDE tem majoritariamente explorado dados de maior granularidade, vindos de sistemas tutores inteligentes, que registram cada passo na resolução de um item ou tarefa, mas que ainda assim necessitam de uma ordenação entre o que deve ser ensinado. Com isso, sistemas inteligentes podem usar o modelo resultante para representar os conceitos de um domínio e como eles são operacionalizados dentro do sistema.

O mapeamento de habilidades para os itens fornece uma maneira de converter um modelo cognitivo simbólico que um especialista possa produzir, em um formato que possa ser usado para processamento estatístico, ao fazer previsões sobre o desempenho do estudante e guia-lo nas atividades de aprendizagem. Como os modelos não são

determinísticos - pois devem incorporar fatores como o acerto casual e o deslize - é necessária a modelagem da incerteza ao inferir o conhecimento a partir de dados de desempenho. Os modelos estatístico e cognitivo são intrinsecamente ligados conceitualmente, ao que se refere a um modelo preditivo de aprendizagem, cada um com seu aprofundamento na literatura.

Este formato de modelo cognitivo é conhecido como modelo de componentes de conhecimento (*knowledge components*) na comunidade de mineração de dados educacionais e como matriz-Q (*Q-matrix*) na psicometria. Tais modelos representam classes de modelos de variáveis latentes que permitem associar a resposta de um aluno à posse de habilidades subjacentes à resolução dos itens avaliativos aplicados ao aluno.

3. Trabalhos relacionados

O processo de estimação da estrutura de pré-requisitos a partir de variáveis observáveis, como itens de avaliações, tem sido investigado por muitos pesquisadores [Desmarais, Maluf e Liu, 1995; Hwang, 2003; Desmarais, Meshkinfam e Gagnon, 2006; Pavlik, 2008; Scheines, Silver e Godin, 2014].

No entanto, a modelagem de habilidades ainda é um problema ainda em aberto, pois o conhecimento é uma variável latente, não podendo ser observada ou medida diretamente. Desmarais, Meshkinfam e Gagnon (2006) propõem o algoritmo POKS (*Partial Order Knowledge Structure*) para aprender automaticamente a dependência de estrutura entre itens a partir de nós observáveis (as respostas aos itens), comparando-o ao algoritmo de redes Bayesianas, apresentando melhor desempenho tanto em poder preditivo quanto eficiência computacional. Em Pavlik *et al.* (2008), o algoritmo POKS é aplicado para se fazer a análise as relações entre habilidades, anotadas para cada problema, a partir das respostas dos alunos. Os resultados foram usados para criar agrupamentos de habilidades redundantes, que apresentavam grande grau de covariância. Com isso, o modelo de domínio foi muito simplificado, auxiliando na determinação de sua estrutura.

Tseng *et al.* (2007) usam mineração de regras de associação para descobrir mapas de conceitos, mas sem considerar a incerteza no processo de transferência de conhecimento do aluno para seu desempenho. Brunskill (2011) deriva a estrutura de conceitos a partir de observações ruidosas usando *log likelihood*, calculado entre o modelo de pré-requisitos e o modelo *flat*, em que todas as habilidades são independentes, de modo a estimar qual modelo melhor se ajusta aos dados dos alunos, com resultados demonstrando a melhor aderência do modelo proposto. Scheines, Silver e Goldin (2011) aplicam algoritmos de descoberta causal para derivar uma estrutura de requisitos ao aplicar testes estatísticos sobre as variáveis latentes. Chan, Willemin e Labat (2015) propõem um método de mineração de regras de associação probabilístico, com os estados de conhecimento probabilísticos estimados por um modelo de evidência, de modo a se encontrar uma estrutura a partir dos dados de desempenho.

Este trabalho busca preencher a lacuna de explorar um método automatizado de exploração de relações de pré-requisitos em avaliações de larga escala, em especial às avaliações aplicadas no Brasil - o que parece não ter sido explorado na literatura.

4. Materiais e métodos

Para este trabalho, foram usados os metadados do SAEB 2013, da área de conhecimento de Matemática e focado apenas no 9º ano do ensino fundamental. Embora seja um escopo limitado, trata-se de uma amostra significativa, que avalia todas as habilidades propostas na matriz em todo o país, de forma amostral. Trabalhos futuros expandirão este escopo para outros ciclos e outros anos de aplicação. A hipótese é que a estrutura se mantenha, o que será testado futuramente.

Neste conjunto de dados, cada item aplicado é mapeado a uma habilidade da matriz de referência (*matriz-Q*), contendo ao todo 91 itens que trabalham 37 habilidades com ao menos um item. Além disso, existe o registro de cada resposta dada por cada aluno a cada item aplicado a ele (matriz de respondentes e respostas, com mais de 2 milhões de respondentes). O SAEB tem caráter amostral, sendo aplicado em todo o país, mas não em todas as escolas. O objetivo principal desta avaliação é diagnosticar o desenvolvimento da aprendizagem nacional em cima do que foi estabelecido nas matrizes de referência. Ao contrário de outras avaliações como o ENEM, que busca avaliar o indivíduo, o SAEB aplica menos itens aos alunos (26 versus 45), já que o erro de medida para o indivíduo não é tão importante, dado que o objetivo é avaliar o sistema educacional como um todo. Outra característica desta avaliação é o uso dos chamados *blocos incompletos balanceados* (BIBs), em que o total de itens é organizado em diferentes blocos, que por sua vez, são organizados em cadernos e estes são aplicados aos alunos; no total são 7 blocos de 13 itens, sem repetições entre si. Os cadernos de prova são organizados de modo a conterem 2 blocos de 13 itens cada e cada aluno responde a um caderno apenas.

Para estimar as dependências entre as habilidades usando este conjunto de dados foi usado o algoritmo POKS (*Partial Order Knowledge Structure*, Desmarais, 2006), com a implementação de [Desmarais e Bhatnager, 2015]. Foi usado o software R para a implementação do algoritmo POKS. Neste algoritmo são tomados cada par possível de habilidades, testando se existe interação entre elas, por meio das respostas de desempenho medido pelos itens correspondentes a cada habilidade (acertos e erros). Para se assumir que *A* é *pré-requisito de B*, supõe-se que: se o aluno não dominar a habilidade A, muito provavelmente não dominará B também; de modo análogo, se dominar B, muito provavelmente dominará A. Isso testado a partir de 3 regras:

- i) Existir um suporte mínimo para que $P(A=0|B=0) > \text{limiar}$;
- ii) Existir um suporte mínimo para que $P(B=1|A=1) > \text{limiar}$;
- iii) Apresentar significância estatística no teste de qui-quadrado entre ambas as habilidades.

A Figura 1 ilustra o processo completo.

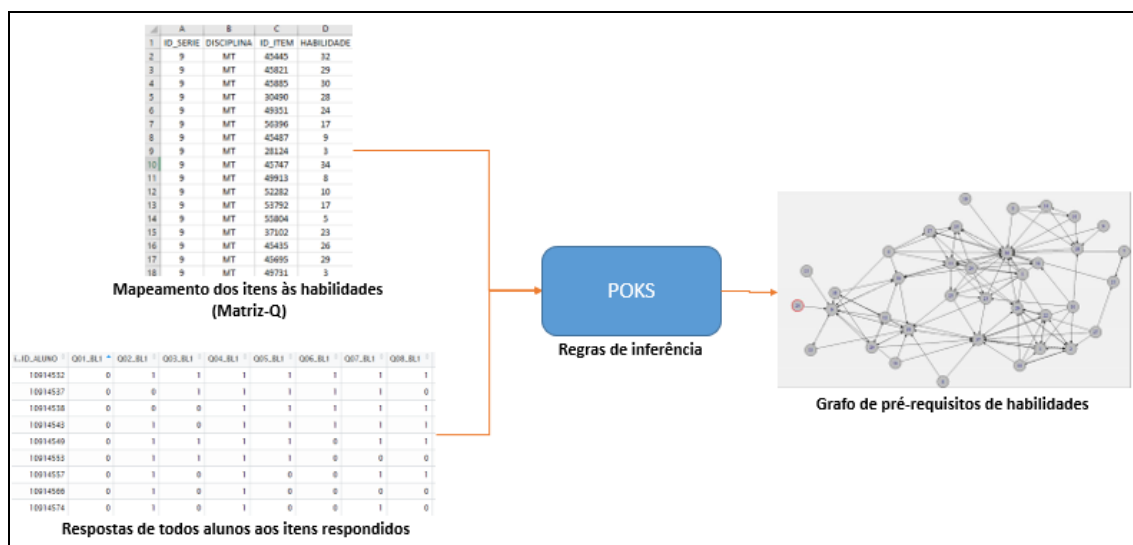


Figura 1. Processo para a estimação do modelo de pré-requisito de habilidades.

O algoritmo traz como parâmetros: o limiar ou probabilidade condicional mínima que testa as regras condicionais entre as habilidades, com um valor de 0,5 na maioria dos casos; o erro alfa para os testes de significância estatística de probabilidade condicional, com valores entre 0,2 e 0,5; o erro alfa para tolerância para o teste de interação (qui-quadrado) entre duas habilidades. A variação destes parâmetros influencia o rigor com que as regras são geradas – quanto maior os erros alfa e menor o limiar, mais regras tendem a ser descobertas. Para este trabalho, foram escolhidos valores para os parâmetros de modo que todo nó (habilidade) apresentasse ao menos uma relação de pré-requisito com outra habilidade, o que deverá ser levado em consideração futuramente.

O resultado é um grafo direcionado, com o significado gráfico das relações como ‘depende de’, ou seja, se uma habilidade A tem uma seta apontando para B, dizemos que A depende de B.

5. Resultados parciais e discussão

O algoritmo POKS se mostrou robusto, com suporte à configuração por BIBs, levando poucos minutos para processar os resultados e a exibição, em um computador pessoal, para a amostra completa de 2 milhões de registros de alunos e 37 habilidades. A Figura 2 traz a rede estimada. Cada nó do grafo representa uma habilidade da matriz de referência e o número em seu interior denota o código da respectiva habilidade.

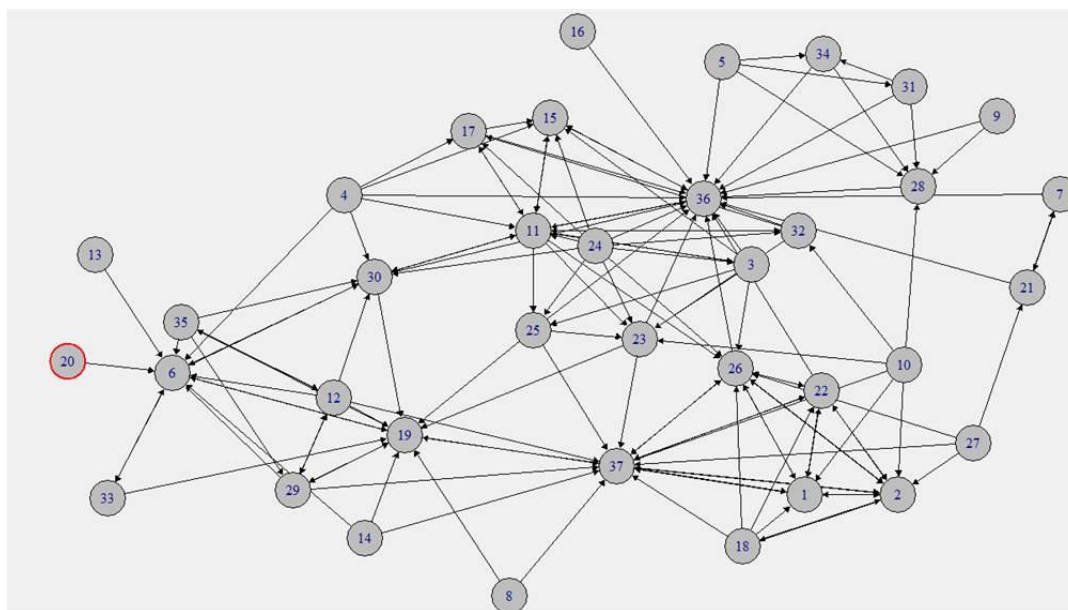


Figure 2. Rede de habilidades e as relações de dependência estimadas método.

Pelo modelo gerado pelo algoritmo e ilustrado Figura 2, podemos notar que algumas habilidades se mostram mais básicas conforme estimado pelo modelo, com um número maior de dependências chegando diretamente até elas e poucas saindo. Destacamos as habilidades 36¹ – *Resolver problema envolvendo informações apresentadas em tabelas e/ou gráficos* e 37 – *Associar informações apresentadas em listas e/ou tabelas simples aos gráficos que as representam e vice-versa*, ambas do tema de tratamento da informação. Já outras se mostram de maneira isolada, como a habilidade 13 - *Resolver problema envolvendo o cálculo de área de figuras planas*, que não é pré-requisito de outras, mas tem como pré-requisito a habilidade 6 - *Reconhecer ângulos como mudança de direção ou giros, identificando ângulos retos e não-retos*.

Este modelo de domínio em matemática, resultante do processo de geração automática das relações de pré-requisitos pode ser usado como componente de um ambiente de ensino inteligente, ao orientar o plano de estudo dos alunos seguindo a ordenação resultante. Assim, um aluno provavelmente começaria pelas habilidades mais básicas estimadas pelo modelo (36 e 37, de tratamento da informação) e disponibilizados os possíveis caminhos conforme as arestas que chegam a estas habilidades, contendo os materiais de estudo para o desenvolvimento destas habilidades.

6. Trabalhos futuros

O resultado da aplicação do método automático pode trazer contribuições importantes para serem exploradas em futuras iterações. Embora as relações tenham trazido relações estatísticas significativas entre os pares de habilidades, é necessário analisar com cautela os resultados. Para isso, deverão ser consultados especialistas no domínio da matemática e em avaliações de larga escala, de modo a identificar possíveis relações espúrias.

¹ Os códigos das habilidades podem ser conferidos na matriz de referência do SAEB, disponível em: <http://portal.inep.gov.br/web/saeb/matrizes-de-referencia-professor>

Alguns relacionamentos são de mão dupla e necessitam de maior investigação (*A depende de B* ao mesmo tempo que *B depende de A*).

Espera-se, com a continuidade dos estudos, aperfeiçoar um método automático para descobrir relações de pré-requisito entre as habilidades que compõem as matrizes de referência na educação básica brasileira e que sejam válidos pedagogicamente, complementando o modelo atual da Teoria de Resposta ao Item. Para isso, o método de estimação precisa ser validado e aprimorado junto a especialistas e avaliadas suas restrições epistemológicas. O método de geração das relações de pré-requisito também deverá conter outras informações, como a dificuldade dos itens, de modo a ponderar melhor a estimação das proficiências das habilidades pelos alunos.

Como produto final, o modelo gerado pode ser usado em ambientes de aprendizagem inteligentes, tanto para fazer um diagnóstico mais preciso da proficiência dos alunos como para auxiliar na criação de sequências didáticas automaticamente, seguindo os caminhos do grafo. Além disso, pode auxiliar a tomada de decisão por gestores educacionais, ao analisar como aprimorar seu sistema de ensino seguindo o modelo diagnóstico gerado.

Referências bibliográficas

- Bruner, J. (1966). *Toward a Theory of Instruction*. Cambridge, MA: Harvard University Press.
- Reigeluth, C., Stein, R. (1983). *Elaboration theory of instruction*. *Instructional-design theories and models*, p. 335-381. Hillsdale, NJ: Erlbaum Associates.
- Vuong, A., Nixon, T., Towle, B. (2011). *A method for finding prerequisites within a curriculum*. *International Conference on Educational Data Mining*, p. 211-216. ISBN: 978-90-386-2537-9
- Nathan, M. J., Koedinger, K. R. (2000). *An Investigation of Teachers' Beliefs of Students' Algebra Development*. *Cognition and Instruction*, 18(2), 209-237.
- Van Patten, J., Chao, C. I., Reigeluth, C. M. (1986). *A review of strategies for sequencing and synthesizing instruction*. *Review of Educational Research*, 56(4), 437-471.
- Clements, D. H., Sarama, J. (2004). *Learning trajectories in mathematics education*. *Mathematical Thinking and Learning*, 6, p. 81-89.
- Koedinger, R. K., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., Rosé, C. P. *Data mining and education* (2015). *WIREs Cogn Sci*, 6:333-353. doi: 10.1002/wes.1350.
- Desmarais, M. C., Maluf, A, Liu, J. (1995). *User-expertise modeling with empirically derived probabilistic implication networks*. *User Modeling and User-Adapted Interaction*, Vol. 5, Issue 3, p. 283-315
- Hwang, G. J. (2003). *A conceptual map model for developing intelligent tutoring systems*. *Computers & Education*. Vol. 40 Issue 3, p. 217 – 235.
- Desmarais, M.C., Meshkinfam, P., Gagnon, M. (2006). *Learned Student Models with Item to Item Knowledge Structures*. *User Modeling and User-adapted Interaction*, 16(5), 403-434.

- Pavlik Jr., P.I., Cen, H., Wu, L., Koedinger, K.R. (2008). Using Item-type Performance Covariance to Improve the Skill Model of an Existing Tutor. In Proceedings of the 1st Intl. Conference on Educational Data Mining, Montreal, Canada, p. 77-86.
- Scheines, R., Silver, E., Goldin, I. (2014). Discovering Prerequisite Relationships among Knowledge Components. In Prerequisites within a Curriculum. In Proceedings of the 4th International Conference on Educational Data Mining, Eindhoven, Netherlands, p. 211-216.
- Tseng, S.S., Sue, P.C., Su, J.M., Weng, J.F., Tsai, W.N. (2007). A New Approach for Constructing the Concept Map. Computers & Education, 49(3), p. 691-707.
- Brunskill, E. (2011). Estimating Prerequisite Structure from Noisy Data. In Proc. of the 4th Intl. Conf. on Educational Data Mining, Eindhoven, Netherlands, p. 217-222.
- Chen, Y., Wullemin, P. H., Labat, J. M. (2015). Discovering prerequisite structure of skills through probabilistic association rules mining. In Proceedings of the 8th Intl. Conference on Educational Data Mining, Montreal, Canadá, p. 77-86.
- Desmarais, M., Bhatnager, S. POKS-skills. Disponível em: <https://github.com/sameerbhatnagar/POKS-skills>. Acessado: 20 Mar 2016.