

Análise Visual de Dados Educacionais: um Estudo de Gênero nos Cursos de Computação da Universidade de Brasília

Luiza A. Hansen, Lucas M. Chagas, Aleteia Patrícia F. de Araújo,
Vinicius R. P. Borges, Maristela Holanda

¹Departamento de Ciência da Computação
Universidade de Brasília (UnB) - Campus Darcy Ribeiro - Brasília, DF - Brasil

luizaahansen@gmail.com, chagas.lucas.mafra@gmail.com,

viniciusrpb@unb.br, aleteia@unb.br, mholanda@unb.br

Abstract. *In the last years, the field of Computer Science has been of little interest for girls attending the undergraduate courses. In the Computer Science Department at the University of Brasília, female students compose less than 10% of the student body. In this paper, we report the pattern of girl students in technology courses by considering three possible conditions to the University: active, terminated and graduated students. For that purpose, that pattern was observed by using a visualization based on Andrews curve.*

Resumo. *Nos últimos anos, a área de Ciência da Computação tem tido pouco interesse das meninas que ingressam na graduação. No Departamento de Ciência da Computação da Universidade de Brasília, o número de meninas é menos que 10% em comparação aos meninos. Neste contexto, este artigo apresenta a visualização de informações sobre alunas dos cursos de computação a partir da sua condição com a Universidade: ainda cursando; desligada ou; já formada. O padrão foi observado utilizando a técnicas de visualização denominada Curva de Andrews.*

1. Introdução

Computação, do latim computatĭo, é um substantivo feminino que tem como definições o ato ou efeito de computar, o conjunto de conhecimentos e técnicas referentes ao uso de computadores e o processamento automático de dados. Na história da computação, pode-se observar a influência de grandes mulheres, como Ada Lovelace, as garotas do ENIAC, a Irmã Mary Kenneth Keller (a primeira norte-americana a receber um doutorado em ciência da computação), entre outras [Othman and Latih 2006] [Agarwal et al. 2016]. Até a década de 80 a computação era uma das áreas com boa participação de mulheres. Diferente dos outros cursos, como Direito e Medicina, a frequência de mulheres na área de tecnologia teve uma grande queda [Cheryan et al. 2013].

Estima-se que o Brasil possua, em média, apenas 18% de concluintes do sexo feminino em diferentes cursos de computação no ano de 2016, de acordo com estatísticas construídas a partir dos dados do INEP, filtrada do censo da educação superior de 2016 realizado pela Sociedade Brasileira de Computação (SBC) [de Computação 2018]. Especificamente, na Universidade de Brasília, os cursos de Ciência da Computação, Engenharia da Computação e Licenciatura da Computação apresentam uma porcentagem muito

baixa da participação das meninas em relação aos meninos, chegando a menos de 10% no ano de 2014.

O processo de descoberta de conhecimento em bases de dados multidimensionais pode auxiliar especialistas em educação na análise de dados educacionais. O emprego de técnicas de visualização pode beneficiar essas tarefas, uma vez que o sistema visual humano identifica e interpreta padrões relevantes em representações gráficas dos dados mais rapidamente em relação às análises descritivas ou exploratórias.

Este artigo objetiva empregar uma técnica de Visualização de Informação, Curva de Andrews, para apoiar a tarefa de mineração de dados educacionais, com a análise do perfil das estudantes concluintes e não-concluintes dos cursos de Computação da Universidade de Brasília.

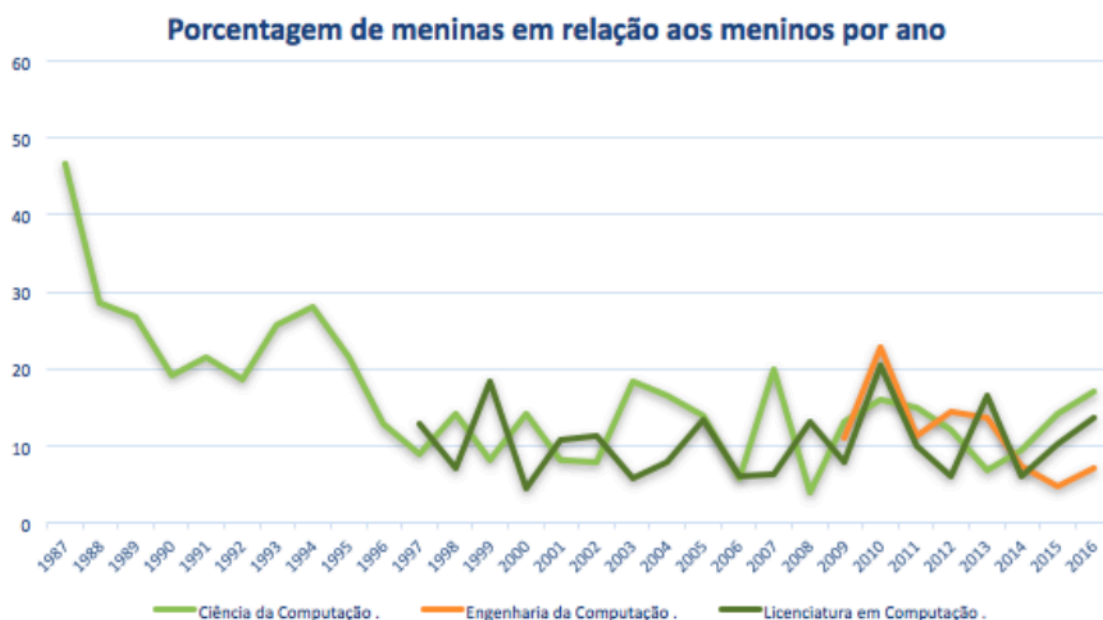


Figura 1. Relação de meninas em relação aos meninos por ano na UnB.

O restante do artigo está dividido nas seguintes seções. Na Seção 2 é detalhado a metodologia seguida no trabalho; na Seção 3 discorre-se sobre os fundamentos teóricos envolvidos neste trabalho; na Seção 4 são apresentados os resultados; por último, na Seção 5 são apresentadas as conclusões e os trabalhos futuros.

2. Metodologia

Com a finalidade de alcançar os objetivos propostos neste artigo, a execução desta pesquisa contempla etapas para o processo de descoberta de conhecimento em bases de dados, conforme ilustrado na Figura 2. A metodologia consiste nas etapas de seleção, pré-processamento, transformação e mineração dos dados, por último, interpretação e análise dos resultados obtidos.

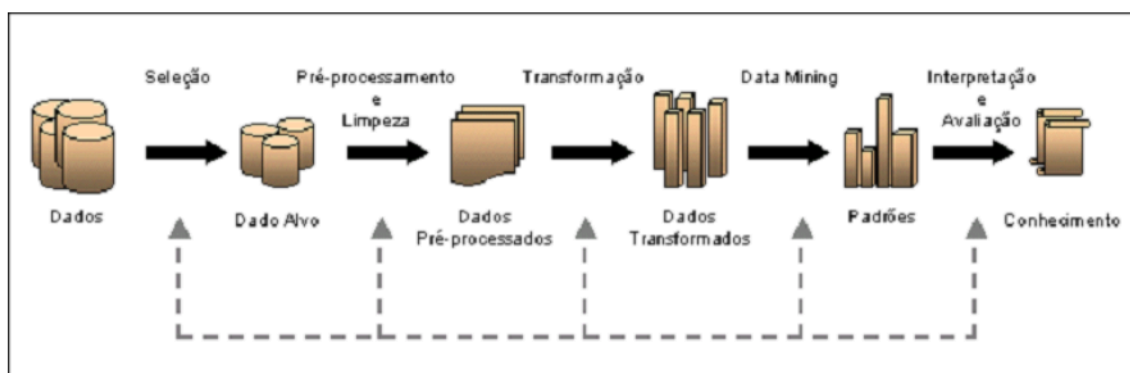


Figura 2. Metodologia para o processo de descoberta do conhecimento.

Os dados foram obtidos do Sistema Acadêmico da Universidade de Brasília e possuem informações sobre alunos dos cursos de Ciência da Computação, Licenciatura da Computação, Engenharia de Computação, Engenharia de Controle de Automação (Engenharia Mecatrônica), Engenharia de Redes de Comunicação e Engenharia de Software.

Os dados coletados dos alunos foram: sexo, data de nascimento, UF de nascimento, se é cotista, o tipo de escola que estudou, a raça, o curso, qual a opção, o período de ingresso na universidade, o período de ingresso na opção, a forma de ingresso na universidade, o período de saída da opção, a forma de saída da opção, o mínimo de créditos para a formatura, o total de créditos cursados. Além disso foram coletados dados sobre as matérias cursadas pelos alunos, ou seja, o número de créditos da matéria, o semestre em que a matéria foi cursada, a nota do aluno, a média do semestre na matéria cursada e quantos créditos o aluno cursou durante o semestre.

Antes de utilizar os dados e fazer a aplicação dos algoritmos explicados na Seção 3 foi necessário adequar os dados para as etapas de mineração de dados, interpolação e avaliação. Foi realizada a normalização dos dados para minimizar os problemas oriundos do uso de unidades e dispersões distintas entre as variáveis. Após o tratamento, os dados foram separados em tabelas e armazenados em um banco de dados relacional.

Em seguida, deve-se especificar as hipóteses sobre os dados em questão e modelar os processos de análise visual e de mineração de dados para respondê-las. Finalmente, devem ser realizados os primeiros experimentos para avaliar e validar as metodologias propostas.

3. Visualização Baseada na Curva de Andrews

Técnicas de visualização multidimensional podem ser utilizadas para apresentar de forma mais intuitiva o conteúdo informacional dos dados, revelando tendências e padrões que propiciam uma melhor tomada de decisão, reduzindo o esforço necessário em tarefas de análise de dados [Garcia-Osorio and Fyfe 2005].

O método de curva de Andrews possibilita uma maneira de visualizar e encontrar uma estrutura em dados multidimensionais. Cada instância da base de dados é definida como uma série de Fourier finita e sua função é colocada em um gráfico para $-\pi < t < \pi$. Assim, cada instância é vista como uma curva entre $-\pi$ e π . Essa fórmula pode ser vista como uma projeção dos dados em um vetor $(\frac{1}{\sqrt{2}}, \sin(t), \cos(t), \sin(2t), \cos(2t), \dots)$.

O objetivo da Curva de Andrews é analisar os pontos de diversos ângulos $-\pi$ a π , o eixo horizontal representa a variação do ângulo, e o eixo vertical, a distância do ponto para o raio, durante a variação. Por isso é possível observar que na Figura 3 as várias linhas representam um ponto de análise. Já as cores representam cada um dos grupos separados previamente (ativas, saíram do curso e formadas).

4. Resultados

A partir dos dados tratados foi possível realizar uma visualização, conforme demonstrado na metodologia proposta na Seção 3. Foram comparados os dados de alunas que 1) estão ativas, ou seja, ainda estão no curso; 2) saíram do curso por algum motivo, dentre eles: desligamento por não cumprir condição (um aluno é considerado "em condição" caso, ao longo de 1 anos não seja aprovado no mínimo de disciplinas necessárias, sendo preciso um plano de estudo para sair dessa condição, caso não consiga cumprir, o aluno é desligado da universidade), desligamento por abandono, desligamento por decisão judicial, desligamento por faltar documentação, desligamento por jubramento, desligamento por força de intercâmbio, desligamento voluntário, desligamento por força de convênio, falecimento, mudança de curso, mudança de turno, novo vestibular, reprovado 3 vezes na mesma disciplina obrigatória, transferência, vestibular para outra habilitação, outros; 3) se formaram [Holanda et al. 2017].

É possível perceber por meio dos gráficos que as linhas 2 (verde) e 3 (vermelho) (as meninas que se formaram e que saíram do curso, respectivamente) são muito semelhantes pois as curvas se sobrepõem no plano visual. As meninas nestes grupos apresentam características semelhantes de acordo com os atributos que as descrevem. No entanto, o perfil das meninas ativas (representadas pela linha 1, azul) diferem das meninas que se formaram ou que saíram do curso, uma vez que não há sobreposição desta com as curvas 2 e 3. Isso significa que é necessário uma análise mais detalhada das meninas nas duas últimas categorias para extrair outros padrões relevantes, o que pode ser feito por meio de um processo de visualização exploratória [Keim 2001] ou empregando-se outras técnicas de visualização mais sofisticadas, como as projeções multidimensionais [Paulovich et al. 2007].

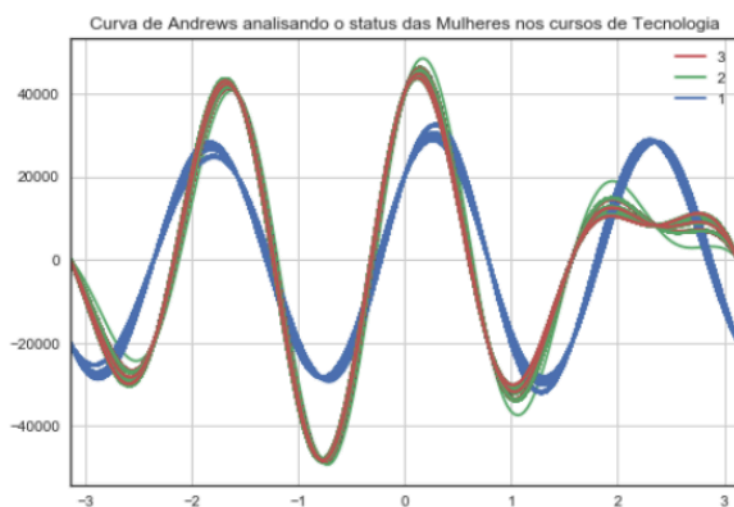


Figura 3. Visualização baseada na Curva de Andrews.

5. Conclusão

Este artigo apresentou um método que utiliza Curva de Andrews para visualizar dados de meninas em cursos de exatas do ensino superior. O método proposto teve como objetivo identificar padrões relevantes pela análise da visualização baseada na Curva de Andrews considerando três perfis de meninas: aquelas se estão ativas, aquelas que se desligaram e aquelas que se formaram. A partir da visualização gerada, pode-se observar, que as meninas que formaram e foram desligadas apresentam características similares de acordo com todos os seus registros acadêmicos, diferentemente das meninas que ainda estão ativas na universidade. Entretanto é necessário uma análise mais detalhada para identificar os fatores que as diferem e quais que são semelhantes.

Como trabalho futuro, além da análise mais aprofundada do perfil das meninas que não são ativas nos cursos, propõe-se uma análise dos dados e suas correlações, como por exemplo a opção do aluno que está fortemente relacionada com o mínimo de créditos para a formatura. Com essa análise será possível excluir alguns dados e ter resultados mais satisfatórios. Como continuação do trabalho será utilizado outras técnicas de visualização de dados, juntamente com um processo de visualização exploratória, para extrair informações relevantes adicionais. Além disso, pretende-se utilizar a visualização como apoio as tarefas de mineração dos dados, podendo prever e analisar o desempenho dos alunos de computação, distinguindo concluintes e não-concluintes.

Referências

- Agarwal, S., Mittal, N., Katyal, R., Sureka, A., and Correa, D. (2016). Women in computer science research: What is the bibliography data telling us? *ACM SIGCAS Computers and Society*, 46(1):7–19.
- Cheryan, S., Plaut, V. C., Handron, C., and Hudson, L. (2013). The stereotypical computer scientist: Gendered media representations as a barrier to inclusion for women. *Sex roles*, 69(1-2):58–71.
- de Computação, S. B. (Accessed 04-Abril-2018). *Educação Superior em Computação Estatísticas – 2016*. <http://www.sbc.org.br/documentos-da-sbc/summary/133-estatisticas/1167-educacao-superior-em-computacao-estatisticas-2016>.
- Garcia-Osorio, C. and Fyfe, C. (2005). Visualization of high-dimensional data via orthogonal curves. *Journal of Universal Computer Science*, 11(11):1806–1819.
- Holanda, M., Dantas, M., Couto, G., Correa, J. M., de Araújo, A. P. F., and Walter, M. E. T. (2017). Perfil das alunas no departamento de computação da universidade de Brasília.
- Keim, D. A. (2001). Visual exploration of large data sets. *Communications of the ACM*, 44(8):38–44.
- Othman, M. and Latih, R. (2006). Women in computer science: no shortage here! *Communications of the ACM*, 49(3):111–114.
- Paulovich, F. V., Oliveira, M. C. F., and Minghim, R. (2007). The projection explorer: A flexible tool for projection-based multidimensional visualization. In *Brazilian Symposium on Computer Graphics and Image Processing*, pages 27–36.