

Redes sociais intra-classe e desempenho acadêmico - uma análise inicial

Luiz Gomes-Jr¹

¹DAINF – UTFPR – Curitiba – PR – Brazil

gomesjr@dainf.ct.utfpr.edu.br

***Resumo.** A compreensão do efeito dos diversos fatores que influenciam o desempenho acadêmico é um requisito importante para a melhoria de práticas educacionais. Como humanos são seres altamente sociáveis, fatores sociais podem ter um papel importante no contexto acadêmico. Este artigo analisa o impacto sobre o desempenho acadêmico de variáveis sociais como o círculo de amigos de um aluno e variáveis de dinâmica de grupo como tendências de agrupamentos. A análise é baseada em dados de 6 turmas da disciplina Bancos de Dados. Este artigo apresenta resultados iniciais que demonstram correlações estatisticamente relevantes entre fatores sociais e desempenho acadêmico.*

1. Introdução

A compreensão de fatores que influenciam o desempenho acadêmico é uma tarefa desafiadora. Existem múltiplas variáveis a considerar e determinar como estas variáveis afetam umas às outras e contribuem para o desempenho acadêmico é um requisito para o desenvolvimento de melhores métodos educacionais.

Um grande desafio em análise de redes sociais é a obtenção dos dados de conexões entre as pessoas. Isto é especialmente desafiador no ambiente acadêmico, uma vez que não há um repositório estabelecido para este tipo de informação. Neste artigo empregamos relacionamentos reportados pelos alunos para a construção das redes sociais de cada turma. Pelo que sabemos, esta é a primeira vez que redes sociais intra-classe são estudadas para identificar influências sobre desempenho acadêmico.

Este artigo analisa o impacto de variáveis sociais sobre o desempenho acadêmico. A análise é baseada em dados de 6 turmas da disciplina Bancos de Dados cursadas por alunos de cursos de computação. Este artigo apresenta resultados iniciais que demonstram correlações estatisticamente relevantes entre fatores sociais e desempenho acadêmico.

2. Trabalhos Relacionados

A falta de dados confiáveis para construir gráficos de redes sociais no contexto acadêmico limita a pesquisa sobre o tema. Uma abordagem para contornar o problema é inferir as relações sociais com base em outras fontes de dados. Yao et al. (Yao et al. 2017) empregaram dados de uso de cartões inteligentes por estudantes nas instalações do campus. A partir dos dados coletados, as relações sociais foram inferidas com base na co-ocorrência de eventos entre os estudantes. Os pesquisadores analisaram dados de vários locais (por exemplo, cantina, biblioteca etc.). Os dados foram usados para construir um modelo de propagação de etiquetas para prever as notas dos alunos com base nas notas de seus pares, alcançando uma precisão de cerca de 40%. Os pesquisadores não apresentam análises sobre medidas de redes sociais e seus impactos no desempenho acadêmico.

Muitas pesquisas investigam o impacto do uso de sites de redes sociais no desempenho acadêmico (ver (Doleck and Lajoie 2018) para uma revisão). O objetivo destes é determinar se o tempo gasto nesses sites pode afetar o desempenho dos alunos. Embora a maioria dos trabalhos mostre que o uso dos sites tem um impacto negativo no desempenho, a associação ainda não é um consenso.

O campo de redes complexas forneceu ferramentas e modelos para a análise de redes sociais em geral (Borgatti et al. 2009). A análise dessas redes é baseada em algoritmos que derivam medições que capturam propriedades do gráfico subjacente. Neste artigo, aplicamos várias medidas para quantificar características sociais de estudantes (nível de nó) e classes (nível de grafo). Por limitações de espaço, direcionamos os leitores para a revisão (da F. Costa et al. 2007) para mais detalhes sobre as medições utilizadas aqui.

3. Análise dos dados

Coleta e limpeza de dados: O conjunto de dados usado para a análise é baseado em dados de seis classes do tópico Bancos de Dados cursadas por estudantes de graduação entre 2016 e 2018. Os alunos são dos cursos de Engenharia de Computação e Sistemas de Informação da Universidade Tecnológica Federal do Paraná (UTFPR).

Coletar dados sociais é uma tarefa desafiadora mas, neste caso, foi simplificada pela natureza do trabalho prático especificado pelo instrutor: construir e analisar a rede social da classe. Para fornecer aos alunos os dados necessários para a tarefa, o instrutor implementou um aplicativo de rede social no qual os alunos devem inserir seus dados no início do período, incluindo suas conexões na turma.

Para cada uma das classes, o grafo de amizades foi criado e somente o maior componente conectado foi retido para representar a rede social da classe. O gráfico foi usado para computar várias medições de rede complexas. As medições foram então integradas com os dados de desempenho do aluno fornecidos pelo instrutor. Os alunos sem dados sociais ou de desempenho (provavelmente desistências) foram excluídos do conjunto de dados. O conjunto de dados final contém 148 estudantes (média de 24.7 alunos/classe, maior classe=33, menor=19).

Variáveis de turma: As variáveis de nível de turma incluídas são: agrupamento médio (avg clustering), comprimento médio do caminho mais curto (avg shortest path len), média de centralidade de intermediação (avg betweenness centr), grau médio (avg degree), assortatividade, eficiência global. As variáveis de nível de classe para as notas são: médias das provas, dos trabalhos, e da nota final.

Os resultados da análise de nível de classe têm menor relevância estatística, pois há menos dados (6 classes) para as inferências. O mapa de calor na Figura 1 mostra a correlação entre as variáveis. As correlações mostram que, em geral, quanto mais conectada é uma classe, melhores são suas notas. A correlação entre o grau médio e a nota média nas provas é de 0,75 ($p = 0,087$), representada na Figura 2. Existem correlações positivas entre outras variáveis que capturam a densidade do grafo, como eficiência global e agrupamento médio. A média do caminho mínimo é correlacionada negativamente com as notas por um motivo semelhante: menores caminhos mais longos aparecem em grafo menos densos. A centralidade de intermediação também é correlacionada negativamente com as notas, o que pode ser devido ao maior número de nós-ponte indicando grupos



Figura 1. Correlações entre as variáveis de classe

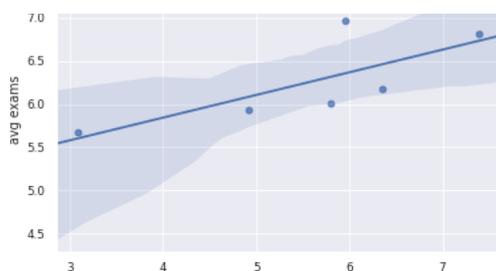


Figura 2. Média de graus (conexões) X média de nota final

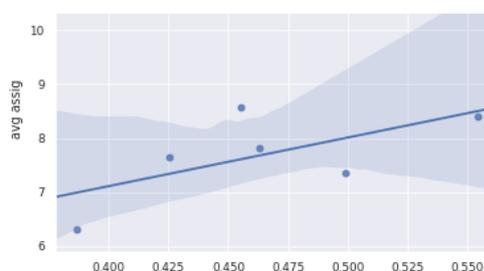


Figura 3. Média de agrupamento X média de nota de trabalhos

isolados no grafo.

Variáveis de alunos: As principais variáveis incluídas no nível de estudante são: grau médio dos vizinhos (average neighbor degree), centralidade de intermediação (betweenness), centralidade de proximidade (closeness), agrupamento (clustering), grau, centralidade de autovetor (eigenvector). Os resultados da análise de nível de aluno têm melhor relevância estatística, pois há mais dados para as inferências.

O mapa de calor na Figura 4 mostra a correlação entre as variáveis no nível de aluno. Existe uma correlação significativa ($p < 0,01$) entre a nota final dos alunos e centralidade de autovetor (correlação: 0,48) e também entre grau médio dos vizinhos (correlação: 0,43), sugerindo que alunos bem conectados tendem a ter notas mais altas. A correlação negativa (fraca) entre as notas e centralidade de intermediação pode ser devido aos estudantes que estão no meio de grupos distintos se sentirem excluídos.

A Figura 5 mostra um gráfico de alunos de acordo com a centralidade do autovetor (normalizado) e nota final. A correlação global entre as variáveis parece pertinente. Alguns padrões interessantes podem ser vistos no gráfico, como uma linha de indivíduos em torno da nota 2, provavelmente indicando alunos com alta carga de créditos. Outro padrão emerge para os alunos com baixa centralidade do autovetor, que são distribuídos regularmente entre as notas. Estes são provavelmente estudantes de outros anos ou cursos que não conheciam muitos dos seus pares. É razoável supor que os fatores sociais não

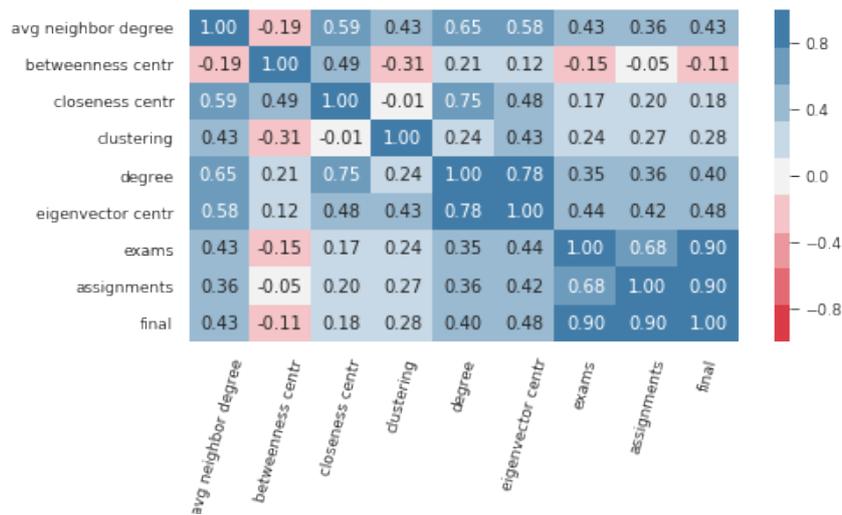


Figura 4. Correlações entre as variáveis de alunos

devem desempenhar um papel importante nesses casos anômalos.

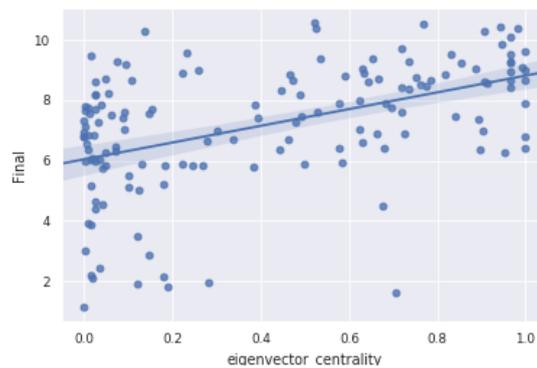


Figura 5. Centralidade de autovetor X nota final

Conclusão: Este artigo demonstra resultados iniciais promissores na análise da associação entre variáveis sociais e desempenho acadêmico. Trabalhos em andamento aprofundarão a análise com a obtenção de mais dados, construção de modelos estatísticos multivariáveis, e uso de técnicas de inferência mais poderosas.

Referências

- [Borgatti et al. 2009] Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916):892–895.
- [da F. Costa et al. 2007] da F. Costa, L., Rodrigues, F. A., Travieso, G., and Boas, P. R. V. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242.
- [Doleck and Lajoie 2018] Doleck, T. and Lajoie, S. P. (2018). Social networking and academic performance: A review. *EAIT*, 23(1):435–465.
- [Yao et al. 2017] Yao, H., Nie, M., Su, H., Xia, H., and Lian, D. (2017). Predicting academic performance via semi-supervised learning with constructed campus social network. In *DASFAA 2017*.