

Investigação Quanto ao Papel de Círculos Sociais no Desempenho Discente no Ensino Superior Utilizando Análise de Redes Complexas

Fernando Concatto
Universidade do Vale do Itajaí
fernandoconcatto@gmail.com

Alex Luciano Roesler Rese
Universidade do Vale do Itajaí
alexrese@univali.br

Rafael de Santiago
Universidade Federal de Santa
Catarina
r.santiago@ufsc.br

Rudimar Luis Scaranto Dazzi
Universidade do Vale do Itajaí
rudimar@univali.br

Anita Maria da Rocha
Fernandes
Universidade do Vale do Itajaí
anita.fernandes@univali.br

ABSTRACT

The individual behavior of human beings is susceptible to influences from their peers. It is known that contact between individuals, both direct and indirect, can foster or inhibit a considerable range of human characteristics and behaviors. This study aims to evaluate the impact of the social context of undergraduate students on their academic performance, understanding "social context" as the average performance of classmates socially close to each student. Our methodology involves reconstructing the underlying social network of a class computationally, using data gathered from a questionnaire applied to the students of the class, and testing the hypothesis that the change in the grade of a student can be accurately modeled as a linear function of the differences between the student's grade and the mean of their peers' grades. The results show that defining a student's social circle as the community they belong to instead of their set of neighbors allows for the construction of statistical models with significantly higher predictive potential.

KEYWORDS

Redes Complexas, Regressão Múltipla, Detecção de Comunidades, Influência Social

1 INTRODUÇÃO

Os seres humanos são caracterizados por serem uma espécie intrinsecamente social, de forma com que o contato entre indivíduos molda, essencialmente, todos os aspectos do cotidiano, tanto de formas positivas quanto negativas, assim como de maneiras sutis e intensas [1]. A estrutura que descreve um conjunto de indivíduos e os relacionamentos entre estes é denominada rede social [2].

Na literatura, observam-se diversos estudos relacionados ao impacto de redes sociais no comportamento de indivíduos. A influência social se faz presente em uma considerável variedade de aspectos, como na disseminação de costumes que favorecem o desenvolvimento de obesidade [3]; na adoção de produtos através do contato com indivíduos-chave, no contexto de marketing [4]; e no uso de drogas, álcool e tabaco por adolescentes [5].

As redes sociais também apresentam elevada importância na educação: Farmer e Rodkin [6] sugerem que certos traços de personalidade em estudantes do ensino fundamental apresentam alta

correlação com seu status social, enquanto Flashman [7] demonstra que estudantes com formas de laços de amizade com colegas que apresentam desempenho acadêmico similar.

Dada a significância das redes sociais no comportamento individual, é possível assumir que estas também moldam a visão dos estudantes acerca da importância do sucesso acadêmico. No trabalho de Blansky et al. [8] sugere-se que, para estudantes do nível 11 em uma escola dos Estados Unidos (equivalente ao Segundo Ano do Ensino Médio no Brasil), a presença de um grupo de amigos com alto desempenho acadêmico tende a fomentar uma melhora nas notas do aluno. Tal resultado se mostra coerente com o trabalho de Flashman [7], além de indicar a existência de um processo de contágio social (fenômeno onde hábitos e comportamentos de um indivíduo se propagam para outros) no contexto das salas de aula evidenciando a capacidade de disseminação de comportamentos em redes sociais. Entretanto, a pesquisa considera que o grupo social que exerce influência sobre um acadêmico é determinado exatamente pelos colegas com os quais o indivíduo se relaciona, desprezando o impacto de características estruturais da rede sob uma perspectiva mais ampla.

O presente trabalho investiga a dinâmica da influência do grupo social em que estudantes estão inseridos sobre seu desempenho acadêmico, efetuando uma comparação entre duas definições de grupo social: (i) a comunidade a qual o indivíduo pertence; e (ii) o conjunto de colegas com os quais o acadêmico possui vínculo social. Tal análise foi realizada aos moldes do estudo relatado por Blansky et al. [8], com a intenção de verificar se o comportamento observado no Ensino Médio dos Estados Unidos também se faz presente em um curso de graduação de uma instituição de ensino superior no Brasil.

Este estudo busca responder à seguinte pergunta de pesquisa: as características dos círculos sociais dos estudantes do ensino superior apresentam influência sobre a tendência geral de seu desempenho acadêmico? Para esta pergunta, as hipóteses nula e alternativa foram definidas da seguinte maneira:

- H_0 : a tendência geral das notas médias dos estudantes do ensino superior não é afetada por círculos sociais.
- H_1 : a tendência geral das notas médias dos estudantes do ensino superior é afetada de forma significativa por círculos sociais.

2 CONTEXTUALIZAÇÃO

Esta seção apresenta a base teórica na qual este trabalho se apoia, detalhando conceitos acerca de redes complexas, tanto em seu contexto psico-sociológico quanto matemático, e explorando o tópico de detecção de comunidades. Conclui-se a seção com um levantamento acerca de fundamentos relacionados à coleta de dados de natureza sociológica, onde apresenta-se uma discussão sobre aspectos positivos e negativos de diferentes metodologias para efetuar tal processo.

2.1 Análise de Redes Complexas

Redes complexas são definidas como grafos cujas topologias apresentam características similares a redes observáveis no mundo real, onde ordem e desordem coexistem [9]. Neste contexto, redes sociais consistem em especializações de redes complexas, onde seus vértices representam atores (indivíduos) e suas arestas representam laços (relacionamentos de qualquer natureza) [2]. Sendo assim, é cabível a aplicação de técnicas e teoremas publicados no contexto de redes complexas na análise de redes sociais, as quais consistem no principal objeto de interesse deste trabalho.

Uma das características mais importantes de redes complexas é o padrão estrutural formado pela organização de suas arestas. No contexto de redes sociais, Christakis e Fowler [3] indicam que o mesmo conjunto de indivíduos pode apresentar grande variabilidade no modo como completa uma tarefa de acordo com a forma com que seus laços sociais estão organizados; desta maneira, uma rede social cuidadosamente estruturada pode ser tão eficaz quanto uma segunda rede com muito mais vértices, porém organizada aleatoriamente.

Sabe-se que, quando há uma relação entre indivíduos, existe também um potencial de influência entre os mesmos. Segundo Simons-Morton e Farhat [5], a influência social representa o efeito que outros causam sobre os comportamentos e atitudes de indivíduos ou grupos. Tal fenômeno também pode causar profundos efeitos sobre os hábitos de estudantes, especialmente devido à grande quantidade de tempo que estes passam interagindo com seus colegas de classe [10]. Na literatura científica, tem-se um considerável arcabouço de estudos que investigam padrões de características de propagação de influência e de que forma seus efeitos se manifestam nos estudantes, buscando correlações entre seu contexto social e outras variáveis de interesse [6–8, 11–13].

Adicionalmente, a dinâmica de influência entre indivíduos pode ser modificada em decorrência das características estruturais da rede subjacente, como a existência de subgrupos coesos de elementos – denominados comunidades. Alguns trabalhos, como o de Huang e Li [14], indicam que a capacidade de propagação de um agente infeccioso abstrato é maior entre indivíduos que pertencem à mesma comunidade (intra-comunidade); no entanto, a própria presença de comunidades na rede tende a limitar a incidência do contágio, pois a transmissão do agente entre indivíduos que pertencem a comunidades diferentes (inter-comunidade) apresenta probabilidade bastante reduzida.

2.1.1 Detecção de Comunidades. A identificação das comunidades que compõem uma rede provê a analistas uma visão mais sofisticada acerca de sua estrutura, permitindo que a mesma seja analisada em termos de seus módulos. O estudo e a detecção de comunidades é

de interesse prático em diversos domínios da ciência – exemplos incluem a análise de páginas da *web*, onde comunidades podem representar páginas que tratam de assuntos similares; a identificação de clientes com interesses similares, habilitando o desenvolvimento de sistemas de recomendação mais eficazes; o mapeamento de reações metabólicas e complexos de proteínas; o reconhecimento de indivíduos em redes criminosas; a identificação de componentes funcionais utilizados na percepção olfativa; entre outros [9, 15, 16].

A otimização da função de Modularidade, desenvolvida por Newman e Girvan [17], é considerada a forma mais popular de efetuar o processo de detecção de comunidades por Fortunato [9]. A função é baseada no conceito da ausência da estrutura de comunidades em grafos aleatórios; sendo assim, define-se um “modelo nulo”, que consiste em uma cópia da rede original porém com suas ligações redefinidas, buscando eliminar a estrutura de comunidades.

Após o estabelecimento do modelo nulo, computa-se a diferença entre a rede original, com seus vértices agrupados de alguma maneira, e seu modelo nulo. Caso não haja diferença, considera-se que o agrupamento dos vértices possui baixa qualidade – isto é, não revela a estrutura de comunidades da rede. Por outro lado, se a diferença for elevada, então o agrupamento sob avaliação é tomado como um particionamento de alta qualidade da rede em suas comunidades [9].

O modelo nulo adotado com maior frequência envolve uma redefinição das ligações de forma a manter a distribuição de graus dos vértices da rede. Sendo assim, para um vértice u qualquer, seu grau tende a ser igual ou similar no modelo nulo, porém o conjunto de vértices aos quais u está conectado tende a ser totalmente diferente. Matematicamente, a probabilidade de dois vértices u e v estarem conectados no modelo nulo é dada por $\frac{k_u k_v}{2m}$, onde k_u representa o grau do vértice u e m expressa o número de ligações presentes na rede. Com isso, define-se a função de Modularidade como:

$$Q = \frac{1}{2m} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} \left[\left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j) \right] \quad (1)$$

onde Q representa a função de Modularidade, m representa o número de ligações da rede, A expressa sua matriz de adjacências, k_i simboliza o grau do vértice i e a função $\delta(C_i, C_j)$ possui valor 1 caso os vértices i e j pertençam à mesma comunidade e 0 caso contrário. A função Q também pode ser trivialmente adaptada para grafos dirigidos e ponderados [9].

2.2 Coleta de Dados Sociológicos

Pesquisas que investigam a influência e a relação entre pares, tanto em contextos educacionais quanto em outras situações, frequentemente utilizam questionários, onde cada participante é responsável por fornecer informações sobre seus pares. Em geral, busca-se compreender quão próximos os indivíduos se encontram socialmente, pois este fator impacta no poder de influência. Neste sentido, a escala da métrica pode ser tanto binária, onde apenas os contatos mais próximos são nomeados, quanto de múltiplos níveis, onde o grau de proximidade entre indivíduos é quantificado como um valor inteiro [8, 11, 12, 18–21].

De acordo com Newman [19], a coleta de dados sociológicos através de questionários ou entrevistas, onde os participantes fornecem informações de forma direta e opinativa, é um processo trabalhoso e sujeito a erros. Nestas situações, o tamanho das redes sociais investigadas tende a ser consideravelmente limitado, devido ao aspecto manual da coleta de dados; além disso, as informações fornecidas pelos participantes são invariavelmente subjetivas e, portanto, profundamente sujeitas a vieses. Assim, ocorrências recentes de significância limitada e breve (uma discordância em uma conversa ou o recebimento de um presente, por exemplo) podem causar um impacto maior do que o apropriado sobre a visão de um participante acerca de outro.

No entanto, segundo Newman [19], as limitações envolvendo esta metodologia de coleta de dados são geralmente compreendidas e aceitas pela comunidade científica, haja vista que métodos alternativos (como análise documental e observação direta) tendem a ser aplicáveis somente em contextos específicos. Além disso, o mesmo autor, em outro trabalho, sugere que: (i) a coleta de dados por meio de questionamentos aos próprios participantes é o método mais comum de acúmulo de informações relacionadas a redes sociais; e que (ii) as incertezas e fragilidades associadas a este método se fazem presentes em praticamente todas as pesquisas de cunho social, não sendo exclusivas aos estudos que tratam da análise de redes complexas [22].

Marsden [23] fornece uma revisão acerca do tópico, apresentando diversos pontos de vista e sintetizando trabalhos que seguem a metodologia. O autor, em termos gerais, toma uma perspectiva positiva, comentando que, apesar das incertezas, dados coletados por questionários e entrevistas são indubitavelmente úteis e importantes, desde que suas limitações sejam observadas.

Tratando de métodos para aprimorar a acurácia dos resultados de questionários e entrevistas, sugere-se: (i) comparação com um padrão conhecido, através do uso de dados de catálogos e de pesquisas já realizadas; (ii) análise de reciprocidade, dado que relações mútuas sugerem um padrão e não apenas uma observação ao acaso; e (iii) teste-reteste, onde a pesquisa é realizada em dois pontos no tempo, sendo esperado que a estrutura geral da rede mantenha-se relativamente estável. No entanto, nenhum método é capaz de garantir a confiabilidade total dos resultados [23].

3 METODOLOGIA

Nesta seção, apresenta-se a metodologia proposta para a investigação da pergunta de pesquisa estabelecida no presente estudo, buscando-se analisar, através da aplicação do arcabouço teórico da área de Análise de Redes Complexas, o impacto do círculo social de estudantes do ensino superior sobre seu desempenho acadêmico.

3.1 Instrumento de Pesquisa

Conforme identificado na Seção 2.2, a literatura científica estabelece que a estratégia mais comum para a coleta de dados em pesquisas na área de redes sociais é a aplicação de questionários e entrevistas, onde os próprios indivíduos sob estudo fornecem de forma direta informações sobre seus laços sociais. Desta forma, assim como nos trabalhos de Blansky et al., Gremmen et al. e Rambaran et al. [8, 11, 12], que possuem objetivos bastante similares ao presente

estudo, optou-se por utilizar a metodologia de aplicação de um questionário para realizar a coleta de dados sociológicos.

De forma similar ao trabalho de Blansky et al. [8], o questionário desenvolvido possui uma pergunta referente ao nível de amizade com cada colega, utilizando uma escala de valores inteiros. A escala utilizada neste estudo segue a teoria de zonas de contato social descrita por Berkman [24], que estabelece seis níveis de proximidade, variando de amigos extremamente íntimos a indivíduos conhecidos apenas de rosto. Como os colegas do respondente serão identificados apenas por nome no questionário, considera-se que indivíduos que pertencem à zona mais distante, em geral, não poderão ser reconhecidos, e portanto seu valor na escala de amizade será igual a zero.

Com isso, estabeleceu-se uma escala de valores inteiros de 0 a 5, com cada valor correspondendo a uma zona de contato social. Em relação à construção do texto de cada opção, considerou-se tanto a metodologia de Blansky et al. [8], que utilizou os termos “melhor amigo”, “amigo” e “conhecido”, quanto de Rambaran et al. [12], que empregou uma escala referente a “gostar de bater papo”. Na Tabela 1, apresenta-se a relação entre a escala utilizada e as zonas de contato social de Berkman [24], além do texto das opções associadas a cada nível de amizade.

O questionário foi construído como uma aplicação *web*, onde cada estudante pôde acessá-lo utilizando uma chave de acesso única, responsável por discernir sua identidade. A página inicial do questionário apresenta informações gerais sobre a pesquisa e sobre seus responsáveis, contendo também um Termo de Consentimento, elaborado de acordo com diretrizes éticas estabelecidas na literatura [25] e com as normas da instituição onde o estudo será realizado. O acadêmico deve ler e aceitar o Termo de Consentimento para que seja possível prosseguir para a próxima etapa da pesquisa.

Para adquirir as notas dos acadêmicos que aceitaram participar da pesquisa, realizou-se um diálogo com as autoridades responsáveis da instituição, visando a obtenção de permissão para acesso às suas bases de dados e extração das informações essenciais para que os objetivos do presente estudo pudessem ser alcançados. Entretanto, em meio a incertezas decorrentes da dependência do setor de Tecnologia da Informação da instituição (que, no período em que a pesquisa estava sendo realizada, passava por uma reestruturação interna) para que os dados pertinentes pudessem ser adquiridos, optou-se por solicitar aos próprios participantes da pesquisa que os mesmos fornecessem suas notas.

Para tanto, incluiu-se no questionário um estágio de coleta de históricos escolares, onde instrui-se que o participante acesse o ambiente de *intranet* da instituição, faça o *download* de seu histórico escolar extraoficial e realize o *upload* do documento para o repositório onde os dados deste estudo estão sendo armazenados. Durante todo o processo, foi reforçado ao participante que suas informações seriam utilizadas única e exclusivamente para que os objetivos da pesquisa pudessem ser alcançados.

A única pergunta do questionário, referente ao nível de amizade entre pares de indivíduos, é realizada uma vez para cada colega do respondente, em ordem aleatória. O conjunto de integrantes da pesquisa foi definido antecipadamente a partir de informações coletadas do sistema de *intranet* da instituição (acessível por qualquer aluno); para compô-lo, realizou-se uma operação de união sobre os integrantes de três disciplinas da penúltima fase do curso de Ciência

Tabela 1: Detalhamento do instrumento de pesquisa elaborado

Nível de amizade	Zona de contato social	Texto da opção
0	<i>Extended zone</i> (zona estendida)	Não conheço.
1	<i>Nominal zone</i> (zona nominal)	Sei quem é, mas não conversamos.
2	<i>Effective zone</i> (zona efetiva)	Conhecido(a); conversamos de vez em quando.
3	<i>Intimate zone B</i> (zona íntima B)	Amigo(a). Nos damos bem.
4	<i>Intimate zone A</i> (zona íntima A)	Amigo(a) próximo(a), conversamos com frequência!
5	<i>Personal cell</i> (célula pessoal)	Melhor amigo(a), conversamos quase todo dia!

da Computação ofertado pela instituição, no primeiro semestre do ano de 2019, resultando em uma amostra com $N = 42$. Na Tabela 2, apresenta-se a quantidade de alunos de cada uma das disciplinas selecionadas.

Tabela 2: Integrantes das disciplinas da penúltima fase do curso de Ciência da Computação (2019/1)

Conjunto	Número de alunos
Empreendedorismo	34
Engenharia da Usabilidade	20
Computação Gráfica	25
União dos conjuntos	42

Seguindo a metodologia do trabalho de Blansky et al. [8], definiu-se que o período de análise da evolução das notas seria de um ano; sendo assim, adotou-se um critério de exclusão referente a acadêmicos que ainda não tenham completado a terceira fase do curso no momento de aplicação do questionário, evitando a presença de instâncias onde não há notas médias registradas para o primeiro semestre do ano de 2018.

Os convites para participação na pesquisa foram enviados aos acadêmicos por meio de mensagens de texto instantâneas via telefone celular, para os casos onde tal informação era de conhecimento dos autores deste estudo. Nas situações onde tal procedimento era inviável, o estabelecimento de contato foi realizado via e-mail.

3.2 Detecção de Comunidades

Para efetuar a detecção de comunidades, optou-se por utilizar a metodologia *Fine-tuned Q*, descrita no trabalho de Chen, Kuzmin e Szymanski [26]. Esta decisão foi tomada com base nas seguintes constatações: (i) o método leva em consideração os pesos das arestas do grafo sob análise; como no presente estudo as relações de amizade são representadas por pesos, tal característica é crucial para que o particionamento da rede em comunidades seja realizado de forma apropriada; (ii) a função de modularidade utilizada inclui um parâmetro responsável por definir a granularidade do particionamento; apesar do mesmo não ser contemplado no presente estudo, sua existência abre caminho para uma grande gama de trabalhos futuros; e (iii) há uma implementação em código aberto do algoritmo de detecção de comunidades para a linguagem de programação Python, a qual é amplamente utilizada neste estudo para os processos de análise de dados.

No entanto, a metodologia de detecção de comunidades selecionada demanda que o grafo seja não dirigido, enquanto que o instrumento de coleta de dados sociológicos detalhado na Seção 3.1 produz informações de natureza essencialmente dirigida, pois cada resposta registrada é composta por uma origem (o acadêmico que está respondendo ao questionário) e um destino (um dos colegas de classe do respondente).

Sendo assim, em uma etapa de preparação para a execução do algoritmo de detecção de comunidades, aplica-se a seguinte transformação sobre os dados de vínculos sociais: para cada par de vértices (u, v) , os pesos das ligações $w(u, v)$ e $w(v, u)$ são substituídos pela média aritmética entre seus dois valores, tornando-os idênticos e, assim, permitindo que a rede social seja representada por um grafo não dirigido. Caso não haja uma ligação entre dois pares de vértices quaisquer, considera-se que seu peso é igual a zero.

Após esta conversão, os pesos do grafo passam por um estágio transformação de escala, visando amplificar a importância de vínculos sociais próximos. Com isso, aumenta-se a probabilidade de que dois indivíduos que possuam uma relação de amizade profunda sejam atribuídos à mesma comunidade. Para tanto, utilizou-se uma função exponencial, tendo como base o peso da aresta e como expoente um valor definido empiricamente, baseado em visualizações do grafo particionado produzidas a partir do algoritmo de *layout ForceAtlas2* [27]. Para a rede social abordada neste estudo, definiu-se que expoente será igual a 7; entretanto, observou-se que todas as transformações com expoentes no intervalo de 6 a 9 produziram resultados praticamente idênticos.

3.3 Análise de Dados

Por meio das estratégias de análise de dados descritas nesta seção, busca-se identificar uma correlação significativa entre a evolução das notas médias dos estudantes investigados e o desempenho acadêmico característico de seus círculos sociais, respondendo assim à pergunta de pesquisa elaborada na Seção 1. Para tanto, empregou-se uma metodologia inspirada no trabalho de Blansky et al. [8], onde um estudo de natureza similar foi realizado no ensino médio dos Estados Unidos.

Em essência, esta etapa consiste na aplicação de uma análise de regressão linear múltipla, onde as variáveis independentes são dadas por representações das diferenças de desempenho entre cada acadêmico e seu grupo social, enquanto a variável dependente é expressa pela evolução da média das notas de cada acadêmico ao longo de um ano.

Mais especificamente, o modelo será composto por seis variáveis independentes. A primeira, g , é dada pelo *ranking* do aluno em

sua turma em termos da média de suas notas no segundo semestre de 2018, com valores maiores representando notas mais altas e vice-versa (isto é, o aluno cuja nota média é inferior a todos os seus colegas possuirá um *ranking* igual a 1). As demais variáveis independentes são dadas pela diferença entre o *ranking* médio dos colegas do acadêmico e seu próprio *ranking*, sendo imposto um particionamento com base no peso dos arcos com origem no acadêmico. Sendo assim, considerando os níveis de amizade descritos na Seção 3.1 (descartando o zero, pois não produz arcos), são estabelecidas cinco variáveis x_k , com $k \in [1, 5]$, onde o conjunto de colegas de cada acadêmico u é dado por $C_k = \{v \in V \mid w(u, v) = k\}$, onde V representa o conjunto de vértices (indivíduos) do grafo. Com isso, o valor de cada variável x_k para um acadêmico u é dado por:

$$x_k = \frac{\sum_{c \in C_k} R(c)}{|C|} - R(u) \quad (2)$$

onde $R(u)$ representa o *ranking* do acadêmico u , calculado exatamente como no caso da variável g .

Por fim, a variável dependente do modelo, denominada y , é tomada como a diferença de *ranking* entre o primeiro semestre de 2019 e o primeiro semestre de 2018. Desta maneira, um valor positivo na variável dependente indica que, ao longo de um ano, o acadêmico obteve um aumento de desempenho em comparação com seus colegas, sob a perspectiva da turma como um todo.

Adicionalmente, será avaliada uma definição alternativa para o conjunto C , baseada na comunidade à qual o acadêmico pertence. Esta metodologia fundamenta-se em estudos como o de Huang e Li [14], onde descreve-se que a processos de influência entre indivíduos que participam da mesma comunidade são mais intensos em comparação com situações onde os indivíduos se encontram em comunidades distintas. Neste caso, considera-se uma sétima variável independente, x_0 , pois dois acadêmicos podem pertencer à mesma comunidade sem que haja um vínculo social entre os mesmos, ainda que tal situação seja infrequente. Sendo assim, define-se que $C'_k = \{v \in V \mid G(v) = G(u) \wedge w(u, v) = k\}$, com $G(u)$ expressando o identificador da comunidade à qual o indivíduo u pertence e $k \in [0, 5]$.

4 RESULTADOS

Nesta seção, são apresentados os resultados obtidos a partir da metodologia elaborada na Seção 3. Inicialmente, descreve-se as informações coletadas por meio da aplicação do instrumento de pesquisa, detalhado na Seção 3.1, tanto em termos de vínculos sociais quanto às notas médias dos acadêmicos; além disso, relata-se as etapas de pré-processamento aplicadas sobre os dados adquiridos, incluindo a execução do procedimento de detecção de comunidades na rede social reconstruída.

Posteriormente, são apresentados os resultados referentes à realização da análise de regressão linear múltipla, descrita na Seção 3.3, possibilitando uma comparação entre as duas definições de círculo social avaliadas neste estudo.

4.1 Aplicação do Instrumento

Dos 42 integrantes da amostra utilizada neste estudo, 37 foram convidados a participar da pesquisa por meio de mensagens instantâneas via telefone celular, enquanto os 5 remanescentes foram

contactados por e-mail. Destes, 34 acadêmicos aceitaram participar na pesquisa; na Tabela 3, apresenta-se de maneira detalhada os resultados do processo de convite à participação na pesquisa.

Tabela 3: Estatísticas dos participantes da pesquisa.

Meio de contato	N. alunos	Respondentes	Cobertura
Telefone celular	37	34	91,2
E-mail	5	0	0
Total	42	34	80,1

Percebe-se que a troca de mensagens via telefone celular foi significativamente mais eficiente do que o método alternativo. Compreende-se que este fenômeno se deu devido à existência de um vínculo social mais profundo entre os autores e os indivíduos cujos números de telefone eram conhecidos, aumentando a probabilidade de colaboração espontânea por parte dos acadêmicos convidados.

Entretanto, dos 34 indivíduos que responderam ao questionário, 4 foram enquadrados no critério de exclusão estabelecido na Seção 3.1, e portanto todos os dados coletados referentes aos mesmos foram excluídos, reduzindo o tamanho da amostra de 42 para 38. Visando identificar um padrão entre os 8 indivíduos que optaram por não participar na pesquisa, realizou-se uma análise acerca do grau de entrada ponderado de cada indivíduo, definido como $d_w(u) = \sum_{v \in V} w(v, u)$ para um indivíduo u qualquer. O gráfico resultante desta análise é apresentado na Figura 1.

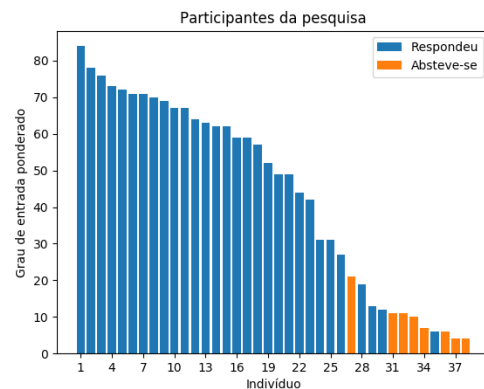


Figura 1: Distribuição do grau de entrada ponderado de cada um dos 38 acadêmicos integrantes da amostra.

Com base na observação de que a maioria dos indivíduos que abstiveram-se da pesquisa possuem um grau de entrada ponderado notavelmente baixo (ou seja, são pouco conhecidos por seus colegas), optou-se por estabelecer um novo critério de exclusão, determinando a remoção dos registros de indivíduos que não responderam à pesquisa e cujos graus de entrada ponderados encontram-se abaixo do primeiro quartil da distribuição (correspondendo aos indivíduos com identificador igual ou maior que 29 na Figura 1). Com isso, o número de acadêmicos analisados foi reduzido para 31, pois o indivíduo de identificador 27 foi mantido na amostra.

Porém, como tal acadêmico absteve-se da pesquisa, não existem informações referentes aos seus vínculos sociais nem às suas notas médias nos dados coletados. Para tratar da ausência de suas notas, optou-se por preenchê-las utilizando a média aritmética das notas dos demais acadêmicos, agrupadas por semestre. Já para os vínculos sociais, realizou-se uma análise de reciprocidade, visando identificar a distribuição discreta de probabilidade do valor de vínculos inversos para cada nível de amizade no intervalo de 0 a 5. Os resultados desta análise são apresentados na Tabela 4, onde as linhas representam o nível de amizade de saída e as colunas expressam a probabilidade de existência de uma relação inversa com a respectiva intensidade.

Tabela 4: Resultados da análise de reciprocidade aplicada sobre os 31 acadêmicos integrantes da amostra.

Nível de amizade	0	1	2	3	4	5
0	0.66	0.18	0.13	0.03	0.00	0.00
1	0.24	0.43	0.25	0.08	0.01	0.00
2	0.12	0.18	0.42	0.26	0.02	0.00
3	0.03	0.07	0.32	0.42	0.15	0.01
4	0.00	0.01	0.06	0.33	0.48	0.12
5	0.00	0.00	0.00	0.04	0.37	0.59

Com isso, foi possível preencher os vínculos sociais com origem no indivíduo de identificador 27 de uma maneira mais assertiva, seguindo os padrões formados pelas relações de reciprocidade entre integrantes da amostra. Além disso, a existência de tal padrão na rede social corrobora a credibilidade dos dados coletados, pois vínculos de amizade são, em grande parte, recíprocos [23].

Por fim, na Figura 2, apresenta-se a rede social da união das turmas analisadas, reconstruída a partir das respostas coletadas. As comunidades detectadas na rede são representadas através das cores atribuídas aos vértices; ou seja, se dois vértices possuem a mesma cor, ambos pertencem à mesma comunidade. A disposição dos vértices foi computada através do algoritmo ForceAtlas2 [27], e a visualização foi produzida por meio do *software* Gephi [28]. O valor de Modularidade atingido foi de $Q = 0.4387$ (um valor mediano [17]), com seis comunidades sendo detectadas.

Optou-se por apresentar a rede em sua forma não dirigida para simplificar a visualização de seus padrões estruturais. Arestas com cor vermelha e maior espessura denotam vínculos sociais íntimos, enquanto arestas amarelas expressam amizades usuais. Vínculos onde os indivíduos se conhecem mas não mantém contato frequente são representados por arestas tendendo à cor azul clara.

4.2 Análise de Regressão

A fim de obter subsídios para responder à pergunta de pesquisa estabelecida neste estudo, foi realizada uma análise de regressão linear múltipla sobre os dados coletados, seguindo a metodologia delineada na Seção 3.3. Adiante, são expostos os resultados obtidos a partir da análise de regressão e da aplicação de testes estatísticos para a verificação de hipóteses de significância dos mesmos.

Após a conversão das notas médias dos acadêmicos para *rankings*, foi realizada uma regressão múltipla entre as variáveis independentes g , x_1 , x_2 , x_3 , x_4 , e x_4 e a variável dependente y , dada pela

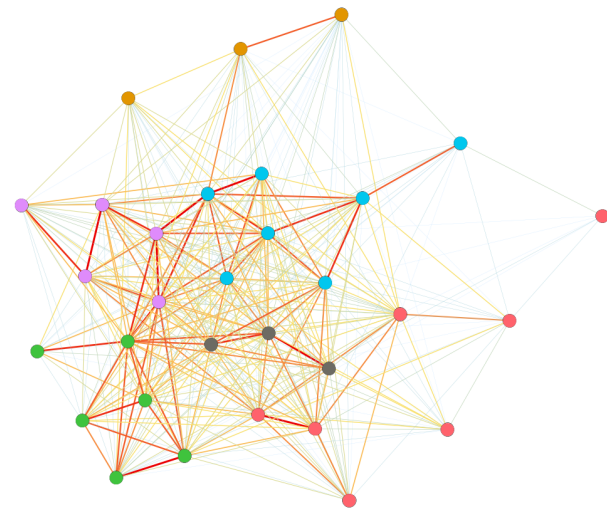


Figura 2: Rede de amizade dos 31 acadêmicos integrantes da amostra, reconstruída a partir dos dados coletados através do questionário.

evolução da nota do acadêmico ao longo de um ano. Neste estágio da análise, foi adotada a primeira definição de grupo social, C_k , onde leva-se em consideração todos os indivíduos com os quais o acadêmico em questão mantém vínculo social, qualquer seja sua intensidade. Na Tabela 5, são apresentados os coeficientes das variáveis independentes, estimados através da aplicação do Método dos Mínimos Quadrados; por brevidade, foi omitido o valor do parâmetro constante do modelo.

Tabela 5: Resultados da regressão linear múltipla aplicada sobre os dados, considerando a definição de círculo social dada em [8].

Variável	Coefficiente	Erro padrão	<i>p</i> -value (teste T)
g	-0.4267	0.756	0.578
x_1	-0.1346	0.446	0.766
x_2	0.4034	0.349	0.259
x_3	-0.3233	0.271	0.244
x_4	-0.0494	0.181	0.788
x_5	0.2672	0.175	0.140

O modelo construído atingiu um coeficiente de determinação de $R^2 = 0.38$; no entanto, como será realizada uma comparação com um modelo constituído por uma variável independente adicional, computou-se também seu coeficiente de determinação ajustado, dado por $R_a^2 = 0.22$. Apesar de tal valor ser razoável para a natureza deste estudo, percebe-se que nenhuma das seis variáveis independentes apresentou significância estatística ($p < 0.05$) quanto à hipótese de que seu respectivo coeficiente é diferente de zero, verificada por um teste *T*. Além disso, por meio da condução de um teste *F* sobre os coeficientes, constatou-se que a hipótese de que todos os coeficientes são iguais a zero não pôde ser rejeitada ($p = 0.054$).

Estes resultados representam uma digressão considerável daqueles relatados no trabalho de Blansky et al. [8], onde apenas uma variável não atingiu um p -value inferior a 0.05, e onde observou-se que a variável g apresentou um coeficiente positivo com um erro padrão extremamente baixo; na regressão realizada neste estudo, um comportamento totalmente oposto foi observado.

Em seguida, foi realizada uma análise similar, porém considerando a segunda definição de círculo social, C'_k , descrita na Seção 3.3. Esta definição leva em consideração as comunidades às quais os acadêmicos pertencem, e implica na necessidade de uma sétima variável independente, x_0 . Os resultados desta análise são apresentados na Tabela 6.

Tabela 6: Resultados da segunda análise de regressão múltipla, considerando a segunda definição de círculo social, a qual favorece a utilização das comunidades detectadas.

Variável	Coefficiente	Erro padrão	p -value (teste T)
g	-1.4468	0.485	0.007
x_0	0.3036	0.232	0.204
x_1	-0.6872	0.284	0.024
x_2	-0.3406	0.156	0.039
x_3	-0.2353	0.153	0.138
x_4	-0.3067	0.158	0.064
x_5	0.1967	0.141	0.178

Neste modelo, obteve-se um coeficiente de determinação de $R^2 = 0.522$ e um coeficiente ajustado de $R^2_a = 0.376$. Um aumento no valor de R^2 já era esperado, devido à inclusão de uma variável independente adicional [29]; no entanto, o ganho de cerca de 67,1% sobre o valor do coeficiente de determinação ajustado indica que a adoção desta definição de círculo social permite explicar a variância dos dados de forma mais eficaz. Para o teste F , obteve-se um p -value de 0.0094, evidenciando a significância do modelo. Além disso, observou-se que três variáveis independentes (g , x_1 e x_2) apresentaram coeficientes significativamente diferentes de zero, obtendo p -values inferiores a 0.05. No entanto, nos três casos, os coeficientes apresentaram valores negativos, contrariando as conclusões obtidas por Blansky et al. [8] em seu estudo.

Para concluir o estágio de análise de dados, foi realizada uma bateria de testes com modelos nulos das redes sociais, onde as seguintes transformações são realizadas: (i) aplica-se uma randomização sobre os destinos de todos os arcos da rede, alterando o conjunto de indivíduos com os quais os acadêmicos mantêm vínculo social; e (ii) os pesos dos arcos são eliminados e gerados novamente de maneira aleatória, seguindo a distribuição de probabilidades original. Com isso, simulam-se situações hipotéticas onde os vínculos sociais entre os acadêmicos desenvolveram-se de maneira diferente; porém, mantém-se inalterado o grau de saída de cada indivíduo. Assim, verifica-se a hipótese de que os resultados observados na análise de regressão foram produzidos por mera coincidência, e não pela existência de um comportamento previsível. Os resultados deste teste, onde 500 modelos nulos foram gerados, são apresentados na Figura 3.

No primeiro caso, com círculos sociais dados por C_k , observa-se que tanto o p -value do teste F quanto o coeficiente de determinação

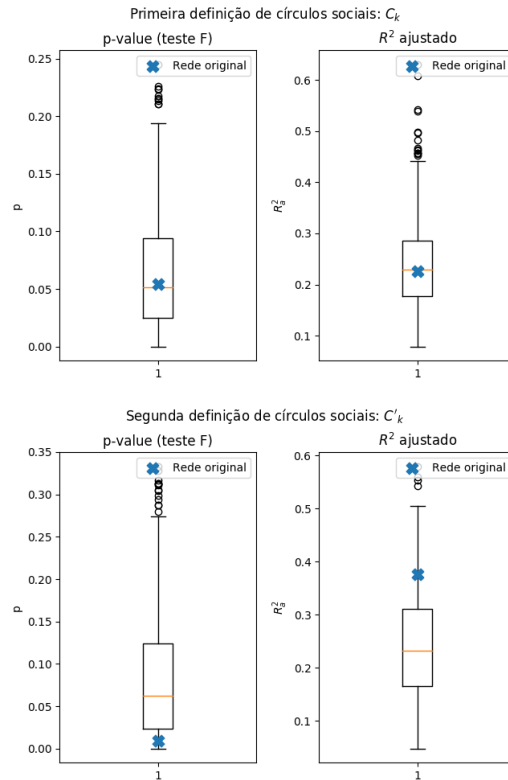


Figura 3: Resultados da aplicação da análise de regressão sobre 500 modelo nulos gerados aleatoriamente.

ajustado da regressão aplicada sobre a rede original mantiveram-se na região da mediana dos modelos nulos, corroborando a ausência de significância estatística observada na Tabela 5.

Ao considerar as comunidades detectadas como entidades representativas dos círculos sociais (C'_k), verificou-se que o p -value da rede original manteve-se próximo da região inferior do espaço gerado pelos modelos nulos, enquanto o valor do coeficiente de determinação manteve-se na região superior. Tal comportamento evidencia a superioridade desta definição de círculos sociais no contexto da análise do impacto da influência social sobre a evolução das notas; no entanto, como ambas as métricas não excederam os limites observados nos 500 modelos nulos avaliados, infere-se que é necessário contemplar variáveis adicionais para explicar o comportamento da evolução das notas dos acadêmicos.

5 CONSIDERAÇÕES FINAIS

No presente trabalho, abordou-se a elaboração e aplicação de um instrumento de pesquisa sobre acadêmicos da nona fase de um curso superior, por meio do qual foram coletados dados referentes às suas notas médias e às relações de amizade entre os mesmos, assim como a condução de uma análise que visou caracterizar o impacto dos círculos sociais dos acadêmicos sobre a evolução de suas notas ao longo de um ano. Relatou-se uma comparação entre

duas definições de círculo social, sendo a primeira referente a vínculos sociais diretos e a segunda referente a comunidades, detectadas computacionalmente através da metodologia *Fine-tuned Q* [26].

A investigação foi realizada por meio da aplicação de uma análise de regressão linear múltipla, apoiada em uma estratégia de validação por modelos nulos. Foi possível constatar que o modelo gerado sobre a primeira definição de círculos sociais produziu resultados sem significância estatística, enquanto a segunda definição permitiu a geração de um modelo com parâmetros significativos, verificados por meio de testes *T* e testes *F*. No entanto, os coeficientes resultantes apresentaram valores negativos, contrariando as observações do trabalho de Blansky et al. [8]. Além disso, levando em conta o comportamento observado na análise por modelos nulos, considera-se que os resultados obtidos não são totalmente conclusivos, impedindo a rejeição ou aceitação da hipótese nula estabelecida no início deste trabalho. Entretanto, constatou-se que, nas circunstâncias do ambiente acadêmico investigado, o uso de comunidades como arquétipos de círculos sociais permite modelar linearmente de forma mais apropriada o impacto causado pelo contexto social sobre o desempenho acadêmico, em comparação com a definição oferecida por Blansky et al. [8].

Para trabalhos futuros nesta linha de pesquisa, sugere-se: (i) realizar a análise sobre conjuntos de turmas de cursos de outras áreas do conhecimento, permitindo uma verificação da generalidade dos resultados observados; (ii) investigar modelos de centralidade, buscando identificar um padrão entre a capacidade de influência e a importância relativa de cada vértice da rede; e (iii) acompanhar os estudantes por um período mais longo, verificando se o comportamento se mantém.

REFERÊNCIAS

- [1] N A Christakis and J H Fowler. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Little, Brown, New York, NY, USA, 2009. ISBN 9780316071345.
- [2] Stanley Wasserman and Katherine Faust. *Social network analysis: methods and applications*. Cambridge University Press, New York, 1994. ISBN 9780521387071.
- [3] Nicholas A. Christakis and James H. Fowler. The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine*, 357(4):370–379, jul 2007. ISSN 0028-4793. doi: 10.1056/NEJMs066082. URL <http://www.nejm.org/doi/abs/10.1056/NEJMs066082>.
- [4] Raghuram Iyengar, Christophe Van den Bulte, and Thomas W. Valente. Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science*, 30(2):195–212, mar 2011. ISSN 0732-2399. doi: 10.1287/mksc.1100.0566. URL <http://pubsonline.informs.org/doi/10.1287/mksc.1100.0566>.
- [5] Bruce G. Simons-Morton and Tilda Farhat. Recent Findings on Peer Group Influences on Adolescent Smoking. *The Journal of Primary Prevention*, 31(4): 191–208, ago 2010. ISSN 0278-095X. doi: 10.1007/s10935-010-0220-x. URL <http://link.springer.com/10.1007/s10935-010-0220-x>.
- [6] Thomas W. Farmer and Philip C. Rodkin. Antisocial and Prosocial Correlates of Classroom Social Positions: The Social Network Centrality Perspective. *Social Development*, 5(2):174–188, jul 1996. ISSN 0961-205X. doi: 10.1111/j.1467-9507.1996.tb00079.x. URL <http://doi.wiley.com/10.1111/j.1467-9507.1996.tb00079.x>.
- [7] Jennifer Flashman. Academic Achievement and Its Impact on Friend Dynamics. *Sociology of Education*, 85(1):61–80, jan 2012. ISSN 0038-0407. doi: 10.1177/0038040711417014. URL <http://journals.sagepub.com/doi/10.1177/0038040711417014>.
- [8] Deanna Blansky, Christina Kavanaugh, Cara Boothroyd, Brianna Benson, Julie Gallagher, John Endress, and Hiroki Sayama. Spread of Academic Success in a High School Social Network. *PLoS ONE*, 8(2):e55944, fev 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0055944. URL <https://dx.plos.org/10.1371/journal.pone.0055944>.
- [9] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5): 75–174, fev 2010. ISSN 03701573. doi: 10.1016/j.physrep.2009.11.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S0370157309002841>.
- [10] Sheretta T. Butler-Barnes, Lorena Estrada-Martinez, Rosa J. Colin, and Brittini D. Jones. School and peer influences on the academic outcomes of African American adolescents. *Journal of Adolescence*, 44:168–181, out 2015. ISSN 01401971. doi: 10.1016/j.adolescence.2015.07.007. URL <https://linkinghub.elsevier.com/retrieve/pii/S0140197115001797>.
- [11] Mariola Claudia Gremmen, Jan Kornelis Dijkstra, Christian Steglich, and René Veenstra. First selection, then influence: Developmental differences in friendship dynamics regarding academic achievement. *Developmental Psychology*, 53(7): 1356–1370, 2017. ISSN 1939-0599. doi: 10.1037/dev0000314. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/dev0000314>.
- [12] J. Ashwin Rambaran, David Schwartz, Daryaneh Badaly, Andrea Hopmeyer, Christian Steglich, and René Veenstra. Academic functioning and peer influences: A short-term longitudinal study of network-behavior dynamics in middle adolescence. *Child Development*, 88(2):523–543, mar 2017. ISSN 14678624. doi: 10.1111/cdev.12611.
- [13] Thomas J. Berndt, Ann E. Laychak, and Keunho Park. Friends' influence on adolescents' academic achievement motivation: An experimental study. *Journal of Educational Psychology*, 82(4):664–670, 1990. ISSN 1939-2176. doi: 10.1037/0022-0663.82.4.664. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0663.82.4.664>.
- [14] Wei Huang and Chunguang Li. Epidemic spreading in scale-free networks with community structure. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(01):P01014–P01014, jan 2007. ISSN 1742-5468. doi: 10.1088/1742-5468/2007/01/P01014. URL <http://stacks.iop.org/1742-5468/2007/i=01/a=P01014>.
- [15] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, jun 2006. ISSN 0027-8424. doi: 10.1073/pnas.0601602103. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0601602103>.
- [16] Rafael Santiago and Luis C. Lamb. Efficient modularity density heuristics for large graphs. *European Journal of Operational Research*, 258(3):844–865, mai 2017. ISSN 03772217. doi: 10.1016/j.ejor.2016.10.033. URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221716308670>.
- [17] M. E J Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 69(2 2), 2004. ISSN 1063651X. doi: 10.1103/PhysRevE.69.026113.
- [18] Amy Bellmore, Xiao Lu Jiang, and Jaana Juvonen. Utilizing Peer Nominations in Middle School: A Longitudinal Comparison Between Complete Classroom-Based and Random List Methods. *Journal of Research on Adolescence*, 20(2): 538–550, mar 2010. ISSN 10508392. doi: 10.1111/j.1532-7795.2010.00640.x. URL <http://doi.wiley.com/10.1111/j.1532-7795.2010.00640.x>.
- [19] M. E. J. Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256, jan 2003. ISSN 0036-1445. doi: 10.1137/S003614450342480. URL <http://epubs.siam.org/doi/10.1137/S003614450342480>.
- [20] Shelley Hymel. Interpretations of Peer Behavior: Affective Bias in Childhood and Adolescence. *Child Development*, 57(2):431, abr 1986. ISSN 00093920. doi: 10.2307/1130599. URL <https://www.jstor.org/stable/1130599?origin=crossref>.
- [21] David Schwartz, Andrea Hopmeyer Gorman, Jonathan Nakamoto, and Tara McKay. Popularity, social acceptance, and aggression in adolescent peer groups: Links with academic performance and school attendance. *Developmental Psychology*, 42(6):1116–1127, 2006. ISSN 1939-0599. doi: 10.1037/0012-1649.42.6.1116. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/0012-1649.42.6.1116>.
- [22] M Newman. *Networks: An Introduction*. Oxford University Press, 2010. ISBN 9780199206650. URL <https://books.google.com.br/books?id=DgTDAQAQBAJ>.
- [23] Peter V. Marsden. Network Data and Measurement. *Annual Review of Sociology*, 16(1):435–463, ago 1990. ISSN 0360-0572. doi: 10.1146/annurev.so.16.080190.002251. URL <http://www.annualreviews.org/doi/10.1146/annurev.so.16.080190.002251>.
- [24] Lisa F. Berkman. Social Support, Social Networks, Social Cohesion and Health. *Social Work in Health Care*, 31(2):3–14, set 2000. ISSN 0098-1389. doi: 10.1300/J010v31n02_02. URL http://www.tandfonline.com/doi/abs/10.1300/J010v31n02_02.
- [25] E.D. de Leeuw, J. J. Hox, and D.A. Dillman. *International Handbook of Survey Methodology*. Routledge, 2008. ISBN 9780203843123. doi: 10.4324/9780203843123. URL <https://www.taylorfrancis.com/books/9780203843123>.
- [26] Mingming Chen, Konstantin Kuzmin, and Boleslaw K. Szymanski. Community Detection via Maximization of Modularity and Its Variants. *IEEE Transactions on Computational Social Systems*, 1(1):46–65, mar 2014. ISSN 2329-924X. doi: 10.1109/TCSS.2014.2307458. URL <http://ieeexplore.ieee.org/document/6785984/>.
- [27] Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE*, 9(6):e98679, jun 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0098679. URL <https://dx.plos.org/10.1371/journal.pone.0098679>.
- [28] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media*, 2009. URL <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>.
- [29] John O. Rawlings, Sastry G. Pantula, and David A. Dickey, editors. *Applied Regression Analysis*. Springer Texts in Statistics. Springer-Verlag, New York, 1998. ISBN 0-387-98454-2. doi: 10.1007/b98890. URL <http://link.springer.com/10.1007/b98890>.