

Lucas Lopes Felipe

Teoria dos Jogos Hedônicos Aplicada ao Problema de Detecção de Comunidade

Rio de Janeiro, Brasil

2021, v-1.0.0

Lucas Lopes Felipe

Teoria dos Jogos Hedônicos Aplicada ao Problema de Detecção de Comunidade

Dissertação de Mestrado submetida à banca de Corpo Docente selecionada pelo Programa de Pós-Graduação em Informática, Instituto de Matemática da Universidade Federal do Rio de Janeiro. Área de Concentração Informática.

Universidade Federal do Rio de Janeiro – UFRJ

Instituto de Matemática

Programa de Pós-Graduação em Informática

Ph.D. Daniel Sadoc Menasché

Rio de Janeiro, Brasil

2021, v-1.0.0

Lucas Lopes Felipe

Teoria dos Jogos Hedônicos Aplicada ao Problema de Detecção de Comunidade/
Lucas Lopes Felipe. – Rio de Janeiro, Brasil, 2021, v-1.0.0-
43p. : il. (algumas color.) ; 30 cm.

Ph.D. Daniel Sadoc Menasché

Dissertação (Mestrado) – Universidade Federal do Rio de Janeiro – UFRJ

Instituto de Matemática

Programa de Pós-Graduação em Informática, 2021, v-1.0.0.

1. Detecção de Comunidades 2. Teoria dos Jogos I. Orientador Ph.D. Daniel Sadoc Menasché. II. Universidade Federal do Rio de Janeiro. III. Programa de Pós-Graduação em Informática. IV. Teoria dos Jogos Hedônicos Aplicada ao Problema de Detecção de Comunidade

Lucas Lopes Felipe

Teoria dos Jogos Hedônicos Aplicada ao Problema de Detecção de Comunidade

Dissertação de Mestrado submetida à banca de Corpo Docente selecionada pelo Programa de Pós-Graduação em Informática, instituto de Matemática da Universidade Federal do Rio de Janeiro. Área de Concentração Informática.

Trabalho aprovado. Rio de Janeiro, Brasil, 27 de Maio de 2021:

Ph.D. Daniel Sadoc Menasché
Orientador - UFRJ/PPGI

Ph.D. Josefino Cabral Melo Lima
UFRJ/PPGI

Ph.D. Daniel Ratton Figueiredo
UFRJ/PESC

Ph.D. Fabricio Murai Ferreira
UFMG/ICEx

Rio de Janeiro, Brasil
2021, v-1.0.0

À você que está lendo.

AGRADECIMENTOS

A vida acadêmica tem sido uma empolgante jornada a medida que quanto mais aprendo mais percebo que há o que ser aprendido, tudo sobre a expectativa do prazer da descoberta por algo que contribua para a rede de conhecimento que nossa civilização vem construindo, e que durante esse processo, estudos que num primeiro momento nasceram de forma desconexa, chegam num estágio de interseção que renova a leitura em cada área, inclusive dando origem a novas.

O trabalho desta dissertação é fruto de uma dessas interseções, unindo um problema do estudo de redes com a teoria dos jogos hedônicos, que me foi cativantemente apresentado pelo meu orientador e amigo Daniel, a quem agradeço por ter sido um mentor nessa jornada, dando as devidas orientações para que eu fosse capaz de fluir com a pesquisa por conta própria, e que assim como o outro Daniel (Shiffman) a quem eu agradei em minha monografia da graduação, me deu motivação e inspiração pela pesquisa. Obrigado Daniel (Sadoc) pelas portas que me abriu e pelos bons diálogos que os jogos hedônicos nos deu, gostaria que tivéssemos mais e sobre outros temas dos tantos incríveis que a Ciência da Computação tem.

Em especial, devo agradecimentos ao Konstantin que me auxiliou enquanto dei seguimento ao trabalho durante meu estágio no INRIA na França. Foi graças a ele que tive a oportunidade de receber a bolsa e ter tido essa ótima experiência. Aproveito também para agradecer a todos os amigos que lá fiz e que me receberam muito bem, como o Giovanni quem me recebeu quando cheguei, e o Guilherme Ricardo, outro aluno da UFRJ, que conheci lá e me auxiliou bastante.

Agradeço os integrantes da banca de minha qualificação, Daniel Ratton e Cabral Lima, que foram fundamentais para dar apoio ao decidir as próximas direções da pesquisa, como a sugestão de recriar um experimento feito pelo Ratton e seu aluno Guilherme Sales.

Agradeço também a todos aqueles com que conversei sobre o Mestrado, por todo apoio, motivação, conselho, encorajamento, opinião, crítica, sugestão e carinho que vocês deram. E por fim, agradeço especialmente você que se interessou em ler esta dissertação, espero que seja uma leitura proveitosa e instigante, e que possa te trazer algum aprendizado e despertar curiosidade por aprender ainda mais.

*“O que há de melhor no homem somente desabrocha
quando se envolve em uma comunidade.”*

Albert Einstein

RESUMO

O problema de detecção de comunidade é muito relevante devido a ampla gama de aplicações que uma solução para ele pode ser útil. Assim como outros problemas analisados pela ótica da ciência da computação, existem dois objetivos fundamentais que uma solução almeja ter: (i) eficiência: referente ao custo para que uma resposta seja dada; (ii) acurácia: quão assertivo é uma resposta comparado com a resposta correta, que é conhecido como *ground-truth*. Neste trabalho estabelecemos uma relação entre teoria dos jogos hedônicos e detecção de comunidades, de modo que visamos ao resolver o primeiro, gerar bons resultados para o segundo. Para isso, apresentemos as contribuições: (i) eficiência: provamos de forma analítica que o algoritmo proposto converge em tempo polinomial; (ii) acurácia: apresentamos um critério de robustez em relação ao parâmetro α que nos auxilia a escolher resposta competitiva. Comparamos o algoritmo proposto com métodos no estado da arte e concluimos que a abordagem dos jogos hedônicos satisfaz as duas métricas, tanto em redes reais quanto sintéticas, quando consideramos a tarefa de rastrear uma comunidade em uma rede dinâmica.

Palavras-chave: detecção de comunidades, teoria dos jogos.

ABSTRACT

Community detection is very relevant due to the wide range of applications that a solution to it can be applied. As like other problems analyzed by the lens of computer science, there are two fundamental goals that a solution aims to have: (i) efficiency: referring to the cost for obtaining an answer; (ii) accuracy: how assertive an answer is compared to the correct answer, known as ground-truth. In this work we establish a relation between hedonic game theory and community detection, so we aim that by solving the former, we would have good results for the latter. For this, we present the contributions: (i) efficiency: we prove analytically that the proposed algorithm converges in polynomial time; (ii) accuracy: we present a criterion of robustness in relation to the parameter α that helps us to choose a competitive answer. We compared the proposed algorithm with state-of-the-art methods, and conclude that the hedonic games approach satisfies the two metrics, both in real and synthetic networks, when considering the task of tracking a community in a dynamic network.

Keywords: community detection, game theory.

LISTA DE ILUSTRAÇÕES

Figura 1 – Robustez de um equilíbrio hedônico.	25
Figura 2 – Rastreamento de comunidade: (a) em redes sintéticas, abaixo do nível de ruído de 0,35, jogos hedônicos são vantajoso para $q/p \leq 0,78$; (b) tolerando níveis de ruído de até $\approx 40\%$	30
Figura 3 – Jogos Hedônicos são eficientes.	30
Figura 4 – Os resultados se estendem a redes reais. A eficiência dos jogos hedônicos está próxima do <i>local improvement</i> , e a acurácia é competitiva contra abordagens de última geração, aproveitando o conhecimento prévio da comunidade rastreada como estado de inicialização.	31
Figura 5 – Fração de nós que não tem incentivo de migrar, para vários valores de α , em redes reais: a linha preta pontilhada representa o particionamento de acordo com o <i>ground-truth</i> , e cada uma das outras linhas corresponde a um particionamento em equilíbrio para um α , $\alpha = 0, 0.1, \dots, 1.0$). Quanto mais reforçada for o tracejado da linha, maior é a acurácia de tal particionamento	33

LISTA DE QUADROS

LISTA DE TABELAS

Tabela 1 – Networks attributes 31

LISTA DE ABREVIATURAS E SIGLAS

CD	Community Detection
HG	Hedonic Games
GT	Ground-Truth
SBM	(Stochastic Block Model)
PBM	(Planted Bisection Model)

LISTA DE SÍMBOLOS

α	Letra grega Alpha, que representa neste trabalho o Parâmetro de Resolução
\in	Pertence

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Motivação	16
1.2	Contribuições	17
1.2.1	Garantia de convergência em tempo polinomial	17
1.2.2	Robustez como critério de seleção	18
1.3	Roteiro	19
2	JOGOS HEDÔNICOS	20
2.1	Terminologia básica	20
2.1.1	A função de α	21
2.1.2	Escolhendo o valor de α	21
2.1.3	Valor Hedônico	21
2.1.4	Potencial	22
2.2	Convergência em tempo polinomial	23
2.2.1	Estabelecendo premissas	23
2.3	Relação entre robustez e acurácia	24
2.3.1	A função acurácia	24
2.3.2	Definindo a robustez	24
2.3.3	Relação através de um exemplo simples	25
3	AVALIAÇÃO DE DESEMPENHO	27
3.1	Algoritmos a serem avaliados	27
3.1.1	Jogos Hedônicos com equilíbrios ‘praticamente’ robustos	27
3.1.2	Spectral Clustering	28
3.1.3	Louvain Method	28
3.1.4	Ensemble Clustering for Graphs (ECG)	28
3.1.5	Local Improvement	29
3.2	Rastreamento de comunidade em redes sintéticas	29
3.3	Rastreamento de comunidade em redes reais	31
3.4	Deteção de comunidade em redes reais	32
3.5	O papel da robustez	33
4	CONCLUSÃO	35
4.1	Revisão	35
4.2	Considerações Finais	36

REFERÊNCIAS	38
APÊNDICES	40
APÊNDICE A – PROVA DO TEOREMA 2.2.1	41
APÊNDICE B – PROVA DO TEOREMA 2.3.1	42
APÊNDICE C – PROVA DO TEOREMA 2.3.2	43

1 INTRODUÇÃO

Neste capítulo será contextualizado o problema de detecção de comunidades. Será apresentada a **motivação**, seguida pelos **objetivos** ao buscar uma solução para o problema de detecção de comunidade usando jogos hedônicos. Concluimos o capítulo listando as **contribuições** trazidas por este trabalho, e descrevendo o **roteiro** do restante da dissertação.

1.1 Motivação

Detecção de comunidade é um dos problemas mais fundamentais em ciência de rede, aprendizado de máquina e análise de rede social (1, 2). Em essência, consiste em encontrar grupos de entidades relacionadas, possivelmente sem conhecimento prévio sobre o *ground-truth* (que doravante será abreviado como GT), que é a informação de qual o grupo que cada elemento pertence (3, 4, 5, 6).

Encontrar comunidades é útil para aplicações em domínios que variam de refinamentos de mecanismos de pesquisa (7), análises de redes sociais de recomendações de anúncios (8), ciência da computação, telecomunicações e bioinformática (9), para citar alguns.

Neste trabalho, focamos no problema de dividir uma rede em dois conjuntos disjuntos de nós, também chamados de bissecção do grafo (*graph bisection*), correspondendo a duas comunidades ou uma comunidade e o restante da rede.

Dentre as questões fundamentais relativas à detecção de comunidade, duas das mais básicas estão relacionadas à eficiência e acurácia:

Eficiência: dado um algoritmo de detecção de comunidade, e uma rede como entrada, *quanto tempo* levará, no pior caso, para o algoritmo convergir em uma solução candidata? – Ou seja, quão demorado é para um algoritmo que se propõe a resolver o problema de detecção de comunidades dar uma resposta.

Acurácia: dada uma solução candidata, quão *próximo* está do *ground-truth*? – Ou seja, quão similar é a resposta dada pelo algoritmo comparado com a resposta correta que é o GT.

Embora as perguntas acima sejam muito básicas, suas respostas ainda estão amplamente abertas.

A maior parte do trabalho em algoritmos de detecção de comunidade, ou é heurística, que funciona bem na prática mas não fornece nenhuma garantia sobre acurácia ou tempo

de convergência (10, 11), ou é passível de análise assintótica em relação à acurácia (12), que normalmente é feita por modelos de rede estilizados como o *Stochastic Block Model* (SBM) (13, 3). O SBM é um modelo gerador de redes sintéticas que apresentam estrutura de comunidade, onde estas são formadas com base em probabilidades definidas por parâmetros no modelo. Em uma série de artigos recentes Hajek *et al.* (14, 5) investigaram algoritmos de convergência polinomial que também se enquadram nesta última categoria.

Neste trabalho, nosso objetivo é analisar um algoritmo de detecção de comunidade que seja analiticamente verificável em relação ao seu tempo de convergência e empiricamente acurado em relação a qualidade de sua resposta. Além disso, pretendemos utilizar uma metodologia que se baseia em princípios básicos, ou seja, na qual a lógica por trás da dinâmica de convergência do algoritmo pode ser interpretada como um jogo e cujas soluções por ele dadas podem ser selecionadas com base em um critério proposto de robustez em relação aos parâmetros do modelo.

Nossa principal intuição para atingir o objetivo mencionado acima é estabelecer uma conexão entre algoritmos de detecção de comunidade e jogos de potencial, conforme estabelecido em (4). Através dessa conexão, pretendemos trazer os resultados promissores da teoria dos jogos de potencial (em específico uma subclasse de Jogos Hedônicos que admitem um potencial, doravante referido como potencial hedônico (4, 15)) para o problema de detecção de comunidade.

Jogos Hedônicos são jogos em que cada vértice de um grafo possui uma função utilidade, que calcula o *valor hedônico* correspondente a ser membro de uma determinada comunidade, que representa o quão satisfeito ele está por fazer parte de tal comunidade. Por uma perspectiva computacional, o valor hedônico dos jogos de potencial hedônico é uma função que pode ser calculada em tempo polinomial, baseando-se apenas em informações locais sobre o número de conexões que um nó possui em uma comunidade, e o número total de nós dessa comunidade. Pela perspectiva de análises de redes sociais, por ser essencialmente local e representar bem as limitações de adquirir e manter as relações sociais, a abordagem do jogo hedônico é particularmente atraente (1).

1.2 Contribuições

Através da conexão entre os jogos de potencial hedônico e detecção da comunidade, fornecemos as seguintes respostas parciais às perguntas acima:

1.2.1 Garantia de convergência em tempo polinomial

Mostramos que o algoritmo de detecção de comunidade baseado em jogos de potencial hedônico converge em tempo polinomial para um *equilíbrio*, ou seja, para uma configuração em que nenhum agente tem incentivo de mudar para outra comunidade, pois

dentre as comunidades existentes, a que é de sua preferência por maximizar a função utilidade resultado no melhor valor hedônico, é a comunidade em que ele já está.

A prova da convergência em tempo polinomial se utiliza das propriedades da função potencial. De fato, o potencial de uma rede (somatório do valor hedônico de todos os nós desta) é restringido por um limite inferior e superior. Então, nós provamos que há um ganho mínimo em potencial para cada vez que um nó move de uma comunidade para outra, que aumenta seu valor hedônico. Aliando as duas observações acima, concluímos que mesmo no pior cenário onde a rede começa numa configuração em que o potencial está no limite inferior, e vai incrementando sempre no menor ganho possível até chegar no limite superior, ainda assim a convergência se dá em tempo polinomial. Os detalhes da prova são apresentados no Capítulo 2 e Apêndice A (Teorema 2.2.1).

1.2.2 Robustez como critério de seleção

Analisamos a convergência para *boas* soluções, ou seja, que estão próximas ao GT. Embora os jogos hedônicos já tenham sido considerados para fins de detecção da comunidade (4), um dos principais desafios para sua aplicação envolve a parametrização do algoritmo. Quando um nó decide em qual comunidade deseja ingressar, um parâmetro que pode variar entre 0 e 1 denotado por α determina o grau de rejeição a membros da mesma comunidades com quem não possui conexão, e seu complemento $(1 - \alpha)$ representa relevância dada aos membros da comunidade que possui conexão. Ou seja, o nó leva em consideração os aspectos positivos (amigos que possui em uma comunidade) e negativos (desconhecidos que há em uma comunidade) quando decide se vale a pena sair da que está para migrar para a que está comparando. O parâmetro α denota o grau de importância desses aspectos positivos e negativos.

Valores menores (resp., maiores) de α correspondem ao aumento da preferência (resp., rejeição) em relação a uma comunidade com mais conhecidos (resp., desconhecidos). Em um extremo, onde $\alpha = 0$, nós não terão nenhum incômodo por ter desconhecidos na comunidade em que pertencem, incentivando a formar o que é conhecido como uma grande coalizão, que é uma única comunidade composta por todos os nós da rede, ou seja, a própria rede. No outro extremo, onde $\alpha = 1$, nós não farão questão de estar em comunidades em que há amigos. O critério de decisão será baseado exclusivamente em evitar desconhecidos, incentivando a se isolarem para ficar numa comunidade com o menor número de desconhecidos possível (mesmo que isso implique em ficar sozinho).

Como pode ser observado, o valor do parâmetro α é crucial para o resultado de convergência do jogo hedônico. Infelizmente, descobrir qual o melhor valor de α que converge num bom resultado não é trivial. Como heurística, em trabalhos anteriores foi atribuído a α a densidade de arestas do grafo (4). Felizmente, descobrimos em investigações numéricas que certos particionamentos correspondem a equilíbrios do jogo potencial para uma ampla

gama de valores de α e que essas soluções consideradas *robustas* são competitivas contra as obtidas por algoritmos no estado da arte (10, 11, 3, 16). Nos Capítulos 2 e 3 e Apêndices B e C exploramos aspectos relacionados à robustez de equilíbrios (Teoremas 2.3.1 e 2.3.2).

1.3 Roteiro

Esta dissertação está organizada na seguinte sequência de capítulos:

1. neste primeiro capítulo fizemos uma introdução ao problema de detecção de comunidades;
2. o Capítulo 2 apresenta o jogo hedônico como proposta de solução ao problema de detecção de comunidade, provando algumas de suas propriedades;
3. o Capítulo 3 compara jogos hedônicos com soluções populares e pontua possíveis próximos passos e extensões;
4. o Capítulo 4 conclui que jogos satisfazem aspectos de eficiência e acurácia e são complementarem a técnicas no estado da arte.

2 JOGOS HEDÔNICOS

Neste capítulo, começamos apresentando a terminologia básica e os pressupostos considerados ao longo do trabalho (Seção 2.1, que é fundamental para entender a dinâmica dos jogos hedônicos). Em seguida, relatamos nossos resultados relacionados à eficiência (Seção 2.2, que diz respeito à convergência em tempo polinomial) e acurácia do algoritmo proposto (Seção 2.3, que estabelece a relação de robustez com acurácia como critério de seleção de equilíbrio). O Teorema 2.2.1 afirma que o algoritmo considerado converge para um equilíbrio em tempo polinomial, e os Teoremas 2.3.1 e 2.3.2 são fundamentais para encontrar *bons* equilíbrios.

2.1 Terminologia básica

Dado um grafo $G = (\mathcal{V}, \mathcal{E})$, com $V = |\mathcal{V}|$ vértices e $E = |\mathcal{E}|$ arestas, uma comunidade é intuitivamente descrita como um conjunto de nós que é densamente conectado entre si, e esparsamente conectado com o resto do sistema. A seguir, tornaremos essa definição de intuitiva para precisa. Para isso, começamos definindo o valor de um nó em uma partição.

Uma comunidade (ou coalizão) é um conjunto \mathcal{S} de vértices, $\mathcal{S} \subseteq \mathcal{V}$. Uma comunidade é dita conectada se para qualquer par de nós em \mathcal{S} há um percurso capaz de partir de um deles e chegar no outro. Um algoritmo de detecção de comunidades particiona a rede em sub-conjuntos $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_K\}$ em que $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$ para $i \neq j$ e $\cup_{i=1}^K \mathcal{S}_i = \mathcal{V}$. Neste trabalho, exceto quando explicitamente indicado de outra forma, assumimos que $K = 2$. Nesse caso, também podemos nos referir ao particionamento como $\{\mathcal{A}, \mathcal{B}\}$ onde \mathcal{A} e \mathcal{B} são as duas comunidades, com $\mathcal{S}_1 = \mathcal{A}$ e $\mathcal{S}_2 = \mathcal{B} = \mathcal{V} \setminus \mathcal{A}$.

Uma comunidade é internamente estável se nenhum nó da comunidade tiver incentivo para deixar a comunidade. Uma comunidade é externamente estável se nenhum nó fora da comunidade tiver incentivo para entrar na comunidade. Uma partição é estável se todas as suas comunidades da rede são interna e externamente estáveis. Neste caso, a configuração é chamada de *equilíbrio*. Para determinar se um nó tem incentivo para migrar de sua comunidade atual, definimos então o valor da relação entre um par de nós, e o valor hedônico de um nó.

Dado um particionamento $\{\mathcal{A}, \mathcal{B}\}$, considere dois nós i e j que estejam na mesma comunidade, ou seja, $i, j \in \mathcal{A}$ ou $i, j \in \mathcal{B}$. Então, o valor do par de nós (i, j) é definido

por (4) como:

$$v_{ij} = \begin{cases} 1 - \alpha, & (i, j) \in \mathcal{E} \\ -\alpha, & (i, j) \notin \mathcal{E} \\ 0, & i = j \end{cases} \quad (2.1)$$

onde $0 \leq \alpha \leq 1$.

2.1.1 A função de α

O parâmetro α controla a *resolução* na qual as comunidades são detectadas. Ele pondera a relevância da presença de links contra a ausência deles ao determinar o valor de um par de nós. Valores pequenos de α favorecem os nós para formar comunidades de granulação grossa com mais links, possivelmente convergindo em uma grande coalizão, que compreende todos os nós em uma única comunidade. Isso ocorre porque quando $\alpha = 0$ nós não incorrem em nenhuma penalidade por terem desconhecidos em suas comunidades. Nesse caso, eles ingressam na comunidade em que possuem mais *conhecidos* (nós com que possua conexão). Quando $\alpha = 1$, em contraste, os nós são penalizados pela ausência de links, mas não são favorecidos por terem conexões, o que resulta em comunidades pequenas de granulação fina. Nesse caso, os nós se juntam à comunidade onde seu número de *desconhecidos* é minimizado.

2.1.2 Escolhendo o valor de α

A escolha de α é crítica para o resultado da detecção da comunidade baseada em jogos hedônicos. Em trabalhos anteriores, os autores consideraram diferentes estratégias para selecionar o *melhor* valor para α (4, 15). Nesta dissertação, em contraste, tomamos uma perspectiva diferente sobre o problema e buscamos equilíbrios que são *robustos* contra mudanças em α . Buscamos, por exemplo, equilíbrios que se mantêm para um grande intervalo de valores de α ou até mesmo para todos $\alpha \in [0, 1]$.

2.1.3 Valor Hedônico

Considere um nó j na comunidade \mathcal{B} . O valor hedônico do nó j é a soma dos valores dos pares de nós (i, j) , para todos $i \in \mathcal{B}$, e é denotado por v_j . Então, a partir de (2.1) temos:

$$v_j = \sum_{(i,j): i \in \mathcal{B}} v_{ij} = d_{Bj} - (n_B - 1)\alpha \quad (2.2)$$

onde d_{Bj} é o grau do nó j apenas contabilizando suas conexões na comunidade \mathcal{B} e n_B é o número de nós na comunidade do nó j , $n_B = |\mathcal{B}|$. Seja \bar{d}_{Bj} o número de ‘estranhos’ a j , ou seja, o número de nós em \mathcal{B} , além de j , menos seu grau nessa comunidade, $\bar{d}_{Bj} = n_B - 1 - d_{Bj}$. Observe que $v_j = d_{Bj}$ (resp., $V_j = -\bar{d}_{Bj}$) se $\alpha = 0$ (resp., $\alpha = 1$).

2.1.4 Potencial

Os jogos hedônicos admitem preferências separáveis aditivamente, ou seja, o valor da comunidade é a soma dos valores dos pares de nós. Logo, o valor (ou *potencial*) de uma comunidade é dado por $\varphi_k = \frac{1}{2} \sum_{i:i \in A_k} v_i$.

O valor de uma configuração (ou *particionamento*) é o somatório do potencial das comunidades existentes, ou seja

$$\Phi = \sum_{k=1}^K \varphi_k = \frac{1}{2} \sum_{k=1}^K \sum_{i,j \in \mathcal{S}_k} v_{ij}.$$

Ao considerar duas comunidades \mathcal{A} e $\mathcal{B} = \mathcal{V} \setminus \mathcal{A}$, $2\Phi = \sum_{i,j \in \mathcal{A}} v_{ij} + \sum_{i,j \in \mathcal{B}} v_{ij}$.

Definição 2.1.1 (Equilíbrio). *Uma solução para o problema de detecção de comunidade é um particionamento em que nenhum nó tem incentivo para entrar (ou sair) para (ou de) qualquer comunidade, ou seja, o valor hedônico de cada nó é maior ou igual ao seu valor hedônico se ele estivesse em outra comunidade.*

Considerando os comentários acima sobre a natureza aditiva do potencial, uma solução para o problema também pode ser caracterizada com base no potencial.

Proposição 1. *Uma solução para o problema de detecção de comunidade é um particionamento em que nenhum nó consegue aumentar seu valor hedônico (e consequentemente o potencial da rede) ao migrar para outra comunidade.*

Algoritmo 1 : Detecção de Comunidades: encontrar o equilíbrio local

Entrada : $G = (\mathcal{V}, \mathcal{E})$, α e particionamento inicial $\{\mathcal{A}, \mathcal{B}\}$
Valores padrões : $\alpha = 2|\mathcal{E}|/(|\mathcal{V}|(|\mathcal{V}| - 1))$, que é a densidade de arestas, e \mathcal{A} e \mathcal{B} de tamanhos iguais, e aleatoriamente separados
Saída : Particionamentos \mathcal{A} and $\mathcal{B} = \mathcal{V} \setminus \mathcal{A}$

```

1 migrou  $\leftarrow true$ 
2 while migrou do
3   migrou  $\leftarrow false$ ;
4   for  $i = 1 \rightarrow |\mathcal{V}|$  do
5     seja  $v$  o  $i$ -ésimo nó de  $\mathcal{V}$ ;
6     seja  $\mathcal{P}$  a coalizão onde  $v$  está,  $\mathcal{P} \subseteq \{\mathcal{A}, \mathcal{B}\}$  e  $\overline{\mathcal{P}} \leftarrow \mathcal{V} \setminus \mathcal{P}$ ;
7     if  $v$  tem incentivo de migrar then
8        $\mathcal{A} \leftarrow \mathcal{P} \setminus \{v\}$  e  $\mathcal{B} \leftarrow \overline{\mathcal{P}} \cup \{v\}$ ;
9       migrou  $\leftarrow true$ ;
10    end
11  end
12 end
13 retorna  $\mathcal{A}$  e  $\mathcal{B}$ 

```

2.2 Convergência em tempo polinomial

Nesta seção, provamos que os jogos hedônicos convergem para um máximo local (equilíbrio) em tempo polinomial. A seguir, apresentamos as ideias chaves da demonstração. A prova completa do Teorema 2.2.1 segue no Apêndice A.

2.2.1 Estabelecendo premissas

Para estabelecer a garantia da convergência em tempo polinomial, consideramos duas premissas:

- i. α é racional, ou seja, $\alpha = b/c$, for $b, c \in \mathbb{N}$
- ii. a cada etapa de um algoritmo eficiente, pelo menos um nó tem um ganho estritamente positivo em seu valor hedônico (consequentemente ao potencial) (ver Algoritmo 1). O algoritmo converge quando não há mais nenhum nó que possa aumentar o potencial, ou seja, quando nenhum nó tem incentivo para se desviar pois todos já estão em sua comunidade de preferência.

Dadas as duas premissas acima, temos o seguinte teorema:

Teorema 2.2.1. *A detecção de comunidade baseada em jogos hedônicos detecta uma comunidade em $O(V^2)$.*

A seguir, descrevemos brevemente os conceitos básicos envolvidos. Observe que o potencial Φ satisfaz $-V^2 \leq \Phi \leq V^2$. Além disso, observe também que o potencial pode ser escrito como $\Phi = U/c$, onde $U, c \in \mathbb{N}$. Agora, considere um nó j que pode ganhar ao migrar entre comunidades. O passo chave na prova consiste em mostrar que tal movimento produz um ganho instantâneo limitado inferiormente por $1/c$. Este limite inferior juntamente com os valores máximo e mínimo de Φ implicam que o algoritmo converge em no máximo $2cV^2$ passos. Esse número é alcançado no pior caso onde na primeira iteração do algoritmo tem-se $U = -V^2q$ e na última iteração $U = V^2q$, e todos os passos dados sempre incrementam $1/c$ ao potencial.

De um ponto de vista teórico, permanece um problema aberto se o teorema acima ainda é válido se α não for um número racional. Do ponto de vista prático, o teorema acima afirma que o algoritmo converge para uma solução em tempo polinomial. A seguir, consideramos o problema da seleção de equilíbrio, ou seja, o problema de encontrar um *bom* equilíbrio que se aproxime bem do GT.

2.3 Relação entre robustez e acurácia

Nesta seção primeiro definimos a função que calcula a acurácia de uma solução dada por um algoritmo em relação ao *ground-truth* (GT). Depois estabelecemos de maneira formal o conceito de robustez. Para por fim, através de um exemplo simples, relacionamos a robustez com acurácia.

2.3.1 A função acurácia

Seja $\mathcal{C}(P, (G = (\mathcal{V}, \mathcal{E}), \sigma))$ a acurácia de uma solução P para o problema de detecção de comunidades (G, σ) , onde G é o grafo e σ o GT. Então

$$\mathcal{C}(P, (G, \sigma)) = \sum_{i \in \mathcal{V}} \mathbb{1}(\sigma_i = P_i) / V \quad (2.3)$$

onde $\mathbb{1}(c)$ é uma função indicadora que é igual a 1 se a condição c for mantida e 0 caso contrário. Observe que qualquer solução para o problema ainda é válida se trocarmos os 0's e 1's correspondentes aos rótulos das duas comunidades. Assume-se que a equação acima é avaliada considerando a alternativa mais favorável para maximizar (2.3).

2.3.2 Definindo a robustez

Nos referimos a um equilíbrio como (α_0, α_1) -robusto se permanece um equilíbrio para todos $\alpha \in [\alpha_0, \alpha_1]$. Nossa intuição é de que se um equilíbrio for robusto para uma ampla faixa de valores de α , é mais provável que ele capture o GT.

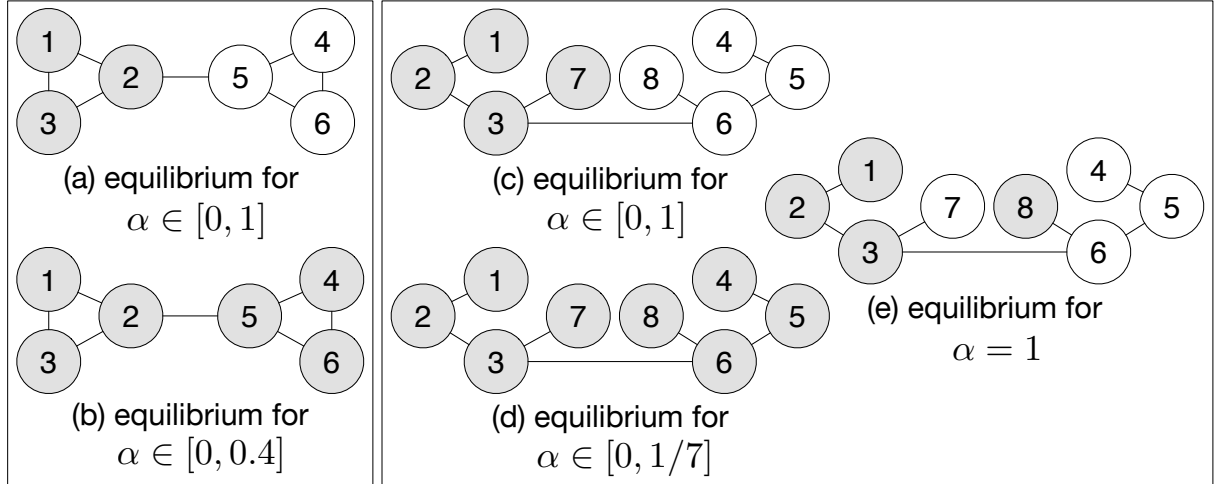


Figura 1 – Robustez de um equilíbrio hedônico.

2.3.3 Relação através de um exemplo simples

Ilustramos a relação entre robustez de equilíbrio e acurácia por meio de dois exemplos simples. Apesar de simples, os exemplos servem para avaliar nossa intuição por trás da razão de usar robustez para a seleção de equilíbrio.

A Figura 1(a) mostra uma rede de 6 nós, composta por dois cliques de 3 nós conectados por meio de uma única aresta.

- As comunidades \mathcal{A} e \mathcal{B} correspondem aos nós $\{1, 2, 3\}$ e $\{4, 5, 6\}$, respectivamente. Se $\alpha = 0$, o valor hedônico de cada nó é igual a 2, que é o número de vizinhos de cada nó em sua comunidade. Se um dos nós 1, 3, 4 ou 6 migra para a outra comunidade, seu valor hedônico diminui para 0. Se os nós 2 ou 5 migram, seu valor hedônico diminui para 1.
- É importante notar que para $\alpha = 0$ a configuração em que todos os nós estão na mesma comunidade (grande coalizão) também é um equilíbrio (Fig. 1(b)). No entanto, esse equilíbrio é menos robusto do que o anterior, mantendo-se apenas para $\alpha \in [0, 0.4]$.

Como outro exemplo, considere a rede nas Figuras 1(c)-(e): o equilíbrio que minimiza o número de pontes entre as comunidades, não sendo a grande coalizão (Figura 1(d)), é o mais robusto (Figura 1(c)).

O teorema a seguir, cuja prova é encontrada no Apêndice B, indica que para mostrar que um dado equilíbrio mantém-se para uma faixa de valores de α , basta considerar os dois extremos da faixa.

Teorema 2.3.1. *Dado um equilíbrio Q que mantém-se para $\alpha = \alpha_0$ e também $\alpha = \alpha_1 > \alpha_0$, então Q é um equilíbrio (α_0, α_1) -robusto.*

Ao procurar por equilíbrios robustos, o seguinte teorema sugere que $\alpha = 0$ é instrumental para encontrar candidatos.

Teorema 2.3.2. *Um equilíbrio para $\alpha = 0$ em que as duas comunidades têm o mesmo tamanho é um equilíbrio para qualquer $\alpha \in [0, 1]$.*

Seja o *desequilíbrio* de um particionamento a diferença entre os tamanhos de suas comunidades. Em um particionamento *perfeitamente equilibrado*, as comunidades têm o mesmo tamanho. A prova do teorema acima é encontrada no Apêndice C, e envolve mostrar equilíbrios perfeitamente equilibrados para $\alpha = 0$, por exemplo, Figs. 1(a) e 1(c), também são equilíbrios para $\alpha = 1$. Então, o resultado segue do Teorema 2.3.1.

Na próximo capítulo mostraremos que o teorema acima é útil para encontrar equilíbrios que capturam o GT com boa acurácia tanto em redes sintéticas como reais.

3 AVALIAÇÃO DE DESEMPENHO

Neste capítulo visamos avaliar o desempenho do algoritmo de jogos hedônicos nas métricas de eficiência (tempo em segundos) e acurácia (conforme definido na Equação 2.3), e comparamos os resultados com os obtidos por outros algoritmos populares na resolução do problema de detecção de comunidades.

A avaliação é feita tanto em redes sintéticas (Seção 3.2), geradas por um modelo adaptado do *Stochastic Block Model* (SBM) para gerar redes de 2 comunidades de mesmo tamanho, como também em redes reais (Seção 3.3), que apresentam essa característica de 2 comunidades majoritárias, mas que não necessariamente sejam de mesmo tamanho.

Através dessas redes sintéticas e reais, e motivados por (17), consideramos o problema de rastrear uma comunidade que evolui ao longo do tempo em uma rede social dinâmica, em que a cada intervalo de tempo, cada nó possui probabilidade f de trocar de comunidade, onde f é aqui referenciado como o *nível de ruído*. Nossos objetivos são:

- i. indicar numericamente a eficiência dos jogos hedônicos para rastreamento e detecção da comunidade;
- ii. estabelecer a relação entre robustez e acurácia;
- iii. ilustrar como a dependência do algoritmo eficiente do jogo hedônico em relação à sua inicialização pode ser aproveitada para fins de rastreamento.

3.1 Algoritmos a serem avaliados

Comparamos a abordagem hedônica com outras quatro estratégias populares:

3.1.1 Jogos Hedônicos com equilíbrios ‘praticamente’ robustos

Para a estratégia utilizando jogos hedônico, nos restringimos a utilizar como resposta *equilíbrios ‘praticamente’ robustos*, que são particionamentos tais que:

- i. nós estão em equilíbrio para $\alpha = 1$;
- ii. pelo menos 90% dos nós não têm incentivos para migrar de comunidade, para todos $\alpha \in [0, 1]$.

Relaxamos o conceito de equilíbrio robusto porque em nossos experimentos verificamos que nem sempre é possível convergir para um equilíbrio onde 100% dos nós

não têm incentivos para se migrar de comunidade, para todos $\alpha \in [0, 1]$. Fica em aberto como trabalho futuro verificar sob quais circunstâncias existe ou não um particionamento robusto para todos os valores de α .

3.1.2 Spectral Clustering

É um método conhecido por ser assintoticamente bom para uma ampla classe de redes (12). Em resumo, o método consiste em calcular um ou mais autovetores do grafo. Em seguida, toma-se a i -ésima componente de cada autovetor como sendo um atributo associado ao nó i . Finalmente, busca-se grupamentos dos nós baseando-se em tais atributos.

3.1.3 Louvain Method

É uma abordagem baseada em modularidade, e empiricamente conhecida por ter um bom desempenho na prática (10). O algoritmo começa com cada nó constituindo sua comunidade. Em seguida, move-se um nó de uma comunidade para outra caso a modularidade associada ao novo grafo seja maior do que a modularidade original. Repete-se o processo até convergência.

A modularidade é definida como

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

onde

- A_{ij} representa o peso da aresta entre os nós i e j ;
- k_i e k_j são a soma dos pesos das arestas atreladas aos nós i e j , respectivamente;
- m é a soma de todos os pesos das arestas do grafo;
- c_i e c_j são as comunidades dos nós; e
- δ é uma função indicadora, $\delta(x, y) = 1$ se $x = y$, e 0 caso contrário.

O termo entre colchetes captura a força das conexões intra-comunidade em comparação com todas as demais conexões. Quanto maior a força dentro de cada comunidade, maior a modularidade da solução.

3.1.4 Ensemble Clustering for Graphs (ECG)

É um método no estado da arte (11) que

- i. gera um conjunto de k particionamentos, por exemplo, usando uma variante do algoritmo *Louvain*;
- ii. produz um novo grafo com os mesmos nós que o original porém as arestas indicam a frequência com que tal par de nós esteve na mesma comunidade dentre todos os k particionamentos gerados na etapa anterior;
- iii. realiza novamente o método Louvain só que dessa vez com o grafo com pesos na arestas gerado na etapa anterior, e retorna como resultado a resposta dada pelo Louvain.

Em nossos experimentos, utilizamos $k = 32$, que é o valor padrão definido em (11).

3.1.5 Local Improvement

É uma estratégia (3) onde dado um particionamento, faz-se o seguinte:

- i. percorra a lista de nós do grafo e sinalize os que possuem mais conhecidos na outra comunidade do que na que está;
- ii. mova todos os nós que foram sinalizados na etapa anterior.

3.2 Rastreamento de comunidade em redes sintéticas

Começamos com redes sintéticas. Para determinar o GT, consideramos o *Planted Bisection Model* (PBM), um membro da classe SBM, que são modelos que primeiro geram as comunidades e depois constrói as arestas da rede a partir de uma probabilidade definida por parâmetro. Um gráfico $G = (\mathcal{V}, \mathcal{E})$ na família de PBM compreende duas comunidades de tamanho n cada, $|\mathcal{V}| = 2n$. Arestas intra (resp., inter) comunidades ocorrem com probabilidade p (resp., q), com $p \geq q$. Ou seja, se um par de nós está em uma mesma comunidade, a probabilidade deles serem conhecidos é p , e caso eles sejam de comunidades distintas, a chance de se conhecerem é q .

Observe que q/p captura a força relativa entre comunidades. Quanto maior a proporção, mais difícil é identificar as comunidades, pois o relacionamento com membros de outra comunidade torna-se tão forte quanto os da mesma comunidade.

Variamos p entre 0,01 e 0,1 em incrementos de 0,01 e ajustamos o valor q de para que um par (p, q) corresponda a uma das relações q/p informadas. Para cada razão q/p , os resultados são obtidos com intervalos de confiança de 95% (menores que 0,1 em todos os casos considerados) a partir de 10 pares (p, q) , onde cada par é avaliado em 25 instâncias distintas de redes de tamanho $n = 500$, cada instância sendo avaliada sob 25 condições iniciais aleatórias.

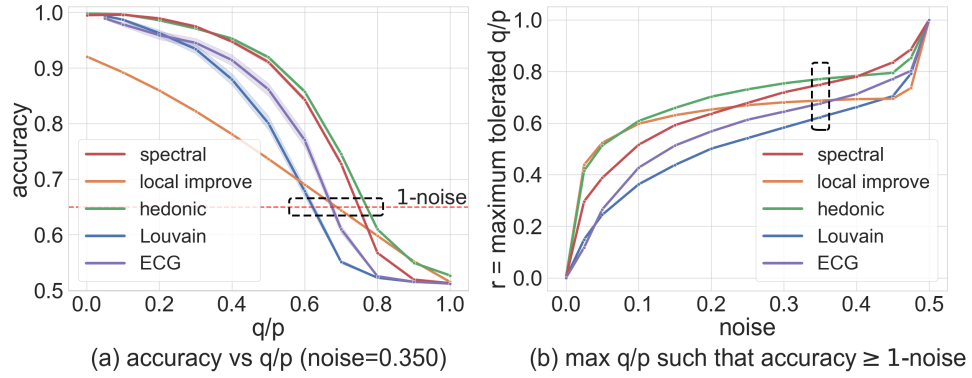


Figura 2 – Rastreamento de comunidade: (a) em redes sintéticas, abaixo do nível de ruído de 0,35, jogos hedônicos são vantajoso para $q/p \leq 0,78$; (b) tolerando níveis de ruído de até $\approx 40\%$.

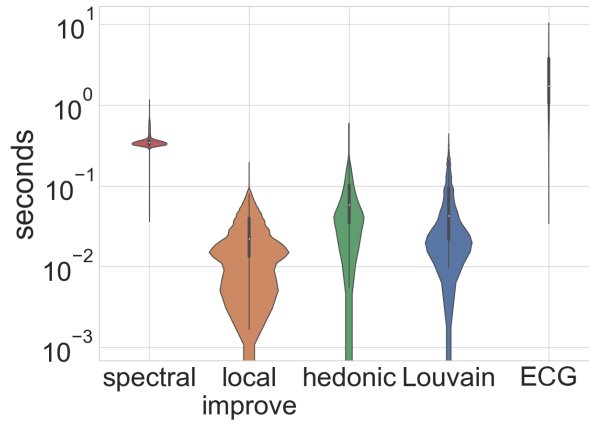


Figura 3 – Jogos Hedônicos são eficientes.

Rastreando comunidades. Nos experimentos a seguir, dado o GT sobre o estado da comunidade em um determinado momento, nosso objetivo é estimar seu estado no momento seguinte. Consideramos $f = 0,35$.

Acurácia. A Figura 2(a) mostra como a acurácia de equilíbrios hedônicos robustos se compara em relação as outras quatro estratégias. Conforme q/p aumenta, a acurácia diminui a medida que o problema se torna mais difícil e a recuperação exata pode ser inviável (3). Para todos os valores de q/p , a detecção da comunidade utilizando a abordagem hedônica permanece mais precisa do que as outras quatro estratégias de detecção de comunidade consideradas. No entanto, em situações mais difíceis ($q/p \geq 0,78$) uma estratégia ingênua que simplesmente repete a entrada na saída supera todos os métodos considerados, pois essa estratégia possui uma precisão de $1 - f$, sendo no caso avaliado $1 - f = 0,65$ que é indicado pela linha vermelha tracejada na Figura 2(a).

Seja r a razão máxima tolerada de q/p tal que para $q/p \geq r$ a estratégia ingênua de repetir a entrada supera suas alternativas. Na Figura 2(a), $r = 0,78$ sob a estratégia

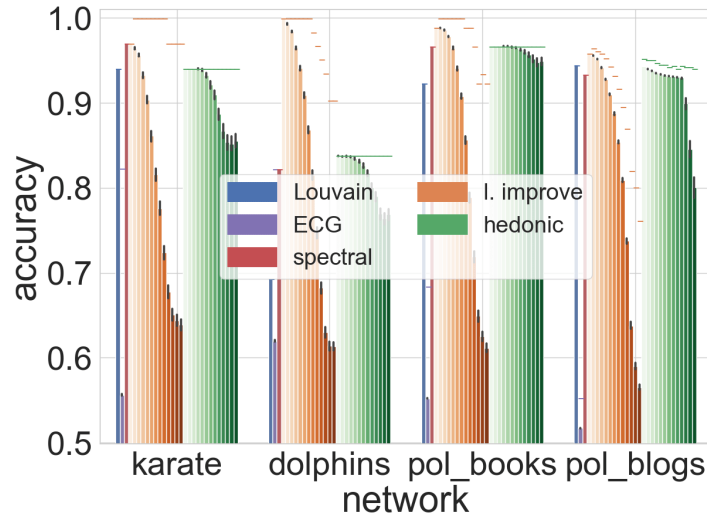


Figura 4 – Os resultados se estendem a redes reais. A eficiência dos jogos hedônicos está próxima do *local improvement*, e a acurácia é competitiva contra abordagens de última geração, aproveitando o conhecimento prévio da comunidade rastreada como estado de inicialização.

Network	V	E	edges density	nodes per community	
Karate	34	78	0.1390	17	17
Dolphins	62	159	0.0840	20	42
Political Blogs	1222	16714	0.0224	586	636
Political Books	92	374	0.0893	43	49

Tabela 1 – Networks attributes

hedônica, e a Figura 2(b) mostra como r varia em função de f .¹ Em particular, a Figura 2(b) sugere que o método hedônico e *spectral* são complementares: enquanto o hedônico é mais eficiente e permite rastrear comunidades até um nível de ruído de $\approx 40\%$, depende-se do *spectral* quando o ruído ultrapassa tal limite.

Eficiência. Figura 3 mostra a distribuição do tempo de convergência das estratégias consideradas. Para tal usamos o *violin plot*. O tempo de convergência do método do jogo hedônico é correspondente ao tempo para encontrar equilíbrios robustos. A Figura 3 indica que *Louvain*, *local improvement* e métodos hedônicos são eficientes, sendo capazes de encontrar equilíbrios em tempo quase linear. Constata-se a baixa complexidade computacional do método hedônico, de acordo com os resultados da Seção 2.2.

3.3 Rastreamento de comunidade em redes reais

Em seguida, avaliamos as abordagens de rastreamento de comunidade, mas agora em quatro redes reais², são elas:

¹ Código-fonte e conjuntos de dados para reproduzir os resultados, além de animações e resultados adicionais disponíveis em <https://github.com/lucaslopes/hedonic/>

² Redes disponíveis em <http://www-personal.umich.edu/~mejn/netdata/>

- i. **political blogs** leva em consideração a polarização política em blogs;
- ii. **political books** é uma rede que indica preferência política em livros;
- iii. **karate** é uma rede formada pela amizade entre 34 membros de um clube de karate que após uma discussão entre os 2 donos, fragmentou-se em 2 comunidades;
- iv. **dolphins** é uma rede social biológica que registra a interação de dois grupos de golfinhos.

A Tabela 1 mostra informações sobre a estrutura de cada rede. O *ground-truth* de cada uma delas é conhecido, e todas compreendem duas comunidades principais. Os resultados da avaliação para redes reais são relatados na Figura 4.

Para cada rede, as estratégias *local improvement* e *hedonic robust* correspondem a 11 barras cada, representando o nível de ruído f variando entre 0 e 0,5 em incrementos de 0,05. O método *spectral*, o algoritmo *Louvain* e o *ECG* correspondem a um apenas uma barra, pois são insensíveis a f já que não permitem alterar o estado inicial da execução do algoritmo. Cada barra mostra a precisão média ao longo de 100 execuções do método correspondente, com intervalos de confiança de 95%, e a acurácia máxima alcançada por cada um dos métodos considerados é marcada em um traço horizontal acima das barras.

Juntas, as Figuras 3 e 4 indicam que a abordagem utilizando o jogo hedônico apresenta uma boa conciliação entre eficiência e precisão. Conforme mostrado na Figura 4, em redes reais os equilíbrios robustos produzidos pela abordagem do jogo hedônico são competitivos contra as quatro alternativas consideradas. Com exceção da rede de golfinhos, a precisão média da abordagem do jogo hedônico está dentro de 0,05 da maior precisão média entre as abordagens consideradas. À medida que o nível de ruído f aumenta, a precisão da abordagem do jogo hedônico diminui, mas essa redução é menor do que a abordagem de melhoria local. De fato, a abordagem do jogo hedônico permanece competitiva na configuração com nível de ruído máximo de $f = 0,5$, que é investigado logo seguir.

3.4 Detecção de comunidade em redes reais

Até então nossa avaliação foi a respeito do problema de rastreamento de comunidade, onde o estado inicial do particionamento é dado por uma versão perturbada do *ground-truth*, com nível de ruído definido por f . Agora, consideramos o problema da detecção da comunidade sem qualquer conhecimento prévio do GT. Consideremos então as mesmas redes reais da seção anterior, e que o algoritmo do jogo hedônico é inicializado com cada nó tendo 50% de chance de fazer parte de cada uma das duas comunidades, que é equivalente a inicializar com GT perturbado em um nível de ruído de $f = 0,5$.

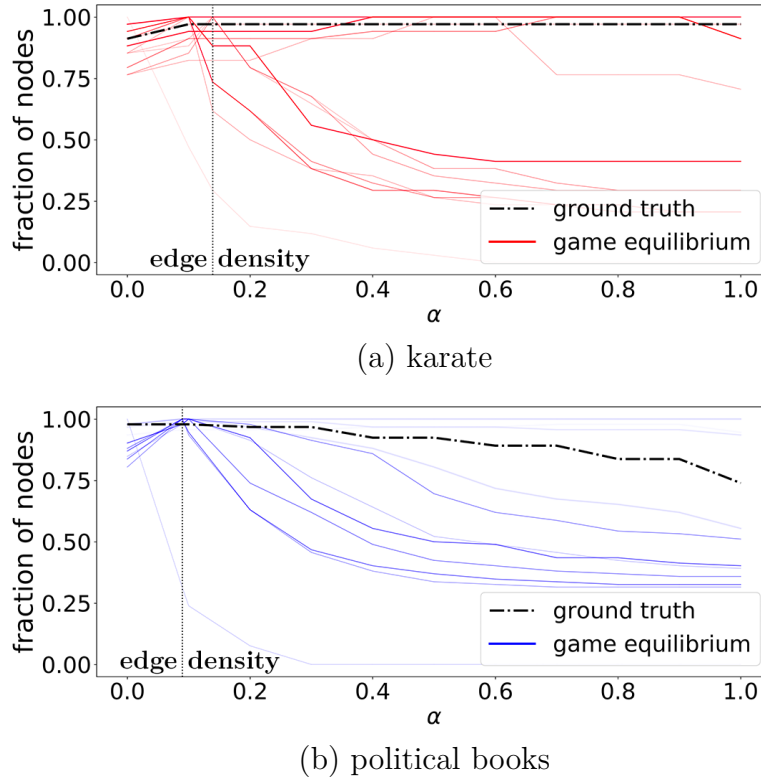


Figura 5 – Fração de nós que não tem incentivo de migrar, para vários valores de α , em redes reais: a linha preta pontilhada representa o particionamento de acordo com o *ground-truth*, e cada uma das outras linhas corresponde a um particionamento em equilíbrio para um α , $\alpha = 0, 0.1, \dots, 1.0$). Quanto mais reforçada for o tracejado da linha, maior é a acurácia de tal particionamento

A densidade de arestas de uma rede é a razão do número de arestas existentes sobre o número possível de arestas, e foi proposta como uma referência para α em (4, 15).

De fato, conforme indicado na análise a seguir, definir α como a densidade de arestas normalmente produz resultados acurados. Esta observação, combinada com a evidência de que os equilíbrios robustos também tendem a ser acurados (Figura 2), motiva a busca por equilíbrios robustos partir de α ter seu valor igual a densidade da aresta da rede.

3.5 O papel da robustez

Nesta seção investigamos a relação entre a robustez de um equilíbrio em relação a α e acurácia. Para tal, encontramos para cada rede um conjunto de equilíbrios para vários valores de α . Para cada um destes, variamos o valor de α e avaliamos como tal alteração afeta a fração de nós que está satisfeita em sua comunidade, ou seja, que não tem incentivo para migrar.

A Figura 5 relata nossos resultados para as redes *karate* e *political books*. Cada

linha na Figura 5 corresponde a um particionamento em equilíbrio. Quanto mais espessa for a linha, maior é a acurácia de tal particionamento. O GT é representada pelas linhas pretas pontilhadas. Para essas redes, observamos que, de fato, o GT tende a ser robusto, de acordo com nossas observações anteriores. Já o GT da redes *dolphins* e *political blogs*, eram menos robustos em relação às mudanças em α do que as outras duas. Em particular, porque o GT nessas redes é mais desequilibrada do que nas redes *karate* e *political books* (ver Tabela 1 e Seção 2.3).

Em resumo, tomando a densidade de arestas como valor de referência para α , pode-se produzir um conjunto de equilíbrios onde a robustez é um critério fundamental para escolher qual desses será a resposta definitiva, especialmente em cenários onde o particionamento é balanceado, ou seja, as comunidades são de tamanhos parecidos.

4 CONCLUSÃO

Neste último capítulo faremos uma revisão e falaremos sobre as conclusões que podemos tirar sobre sua utilidade no cenário de detecção de comunidades

4.1 Revisão

A teoria dos jogos fornece uma interpretação concreta numa perspectiva multiagente para o problema de detecção de comunidade e sua solução (4). Neste trabalho, indicamos que uma determinada classe de jogos de potencial possui um algoritmo eficiente para encontrar equilíbrios. Em seguida, comparamos sua acurácia com soluções populares de última geração, incluindo abordagens baseadas em modularidade (18), métodos de conjunto (11) e métodos espectrais que são conhecidos por serem assintoticamente ideais para uma ampla classe de redes (12).

O método Louvain (10, 18), por exemplo, é uma abordagem eficiente para detecção de comunidade. Nosso método difere da abordagem de Louvain em pelo menos quatro aspectos:

- i. permite pré-determinar o número de comunidades a ser encontrado na rede – lembrando que neste trabalho consideramos apenas duas coalizões, uma formado por uma comunidade de interesse e a outra pelos demais nós da rede;
- ii. é baseado em uma perspectiva multiagente em relação ao problema, pois delega a tomada de decisão de qual nó mover para uma perspectiva local onde cada nó só precisa garantir que esteja em sua comunidade de preferência, que a rede convergirá em um equilíbrio;
- iii. considera a robustez do equilíbrio em relação a um parâmetro de resolução como critério de seleção de boas respostas;
- iv. pode aproveitar do conhecimento prévio sobre as comunidades por meio de sua inicialização de estado, que é instrumental para fins de rastreamento.

Os métodos espectrais produzem comunidades de alta qualidade, mas sua aplicabilidade a problemas de larga escala é dificultada pela complexidade computacional de $O(V^3)$. Em (19) os autores propõem heurísticas para contornar a complexidade dos métodos espectrais. Os jogos hedônicos são complementares a (19), onde no problema de rastreamento da comunidade, os jogos hedônicos podem ser a melhor escolha, pois nos

permitem aproveitar do conhecimento prévio sobre o *ground-truth*. Quando esse conhecimento torna-se desatualizado a ponto de se tornar irrelevante, por exemplo, $f = 0, 5$, os métodos espectrais (19) ou ECG (11) podem ser usados esporadicamente para resgatar o estado da rede.

Mostramos que o método proposto baseado em teoria dos jogos pode encontrar uma comunidade candidata em tempo $O(V^2)$. Além disso, consideramos usar um critério para selecionar ‘bons’ equilíbrios. De fato, dois dos principais problemas clássicos no estudo da teoria dos jogos computacionais são determinar:

- i. complexidade de equilíbrio (20)
- ii. seleção de equilíbrio (21).

A complexidade de equilíbrio refere-se ao problema de determinar a complexidade para encontrar *um* equilíbrio de Nash. Enquanto que a seleção de equilíbrio é o problema de determinar, entre todos os equilíbrios, qual é *o* melhor. Ao estabelecer uma conexão entre esses dois problemas fundamentais da teoria dos jogos, e a detecção da comunidade, visamos a adoção de ferramentas da teoria dos jogos, como por exemplo estratégias evolucionárias de jogos evolutivos, para lidar com os problemas de detecção de comunidade.

Alguns de nossos resultados numéricos foram obtidos usando redes sintéticas e grafos gerados pelo SBM. Alguns resultados de impossibilidade sobre a viabilidade de detecção de comunidade em dados sintéticos foram relatados em (22, 13). Investigar até que ponto podemos melhorar nossos resultados numéricos à luz desses trabalhos também é um tópico interessante para pesquisas futuras.

4.2 Considerações Finais

Detecção de comunidade é uma peça fundamental no *pipeline* atual de ciência de dados. Neste artigo, investigamos a eficiência e a acurácia de um algoritmo de detecção de comunidade baseado em jogos hedônicos. Primeiro, mostramos que o algoritmo considerado é comprovadamente eficiente: seu tempo de computação polinomial corresponde à complexidade de encontrar um equilíbrio de *Nash*. Em seguida, indicamos sua capacidade de rastrear e detectar comunidades de forma acurada.

Em particular, o resultado produzido pelo método proposto pode aproveitar do conhecimento prévio de como estavam as comunidades, utilizando-a como estado inicial para a partir deste encontrar um equilíbrio local, o que mostrou-se instrumental para fins de rastreamento.

Ao introduzir o novo conceito de robustez de equilíbrio em relação a um parâmetro de resolução, mostramos que o rastreamento da comunidade com base em jogos hedônicos supera as soluções no estado da arte (4, 15).

Esperamos que este trabalho abra uma série de caminhos interessantes para investigações futuras, na fronteira entre a teoria dos jogos e detecção de comunidade, incluindo extensões para detectar múltiplas comunidades, cada uma correspondendo a um máximo local da função potencial, e a comparação entre equilíbrios robustos e soluções de máximo *likelihood* (3, 15, 2) também é outro interessante tópico para trabalhos futuros.

REFERÊNCIAS

- 1 SHEIKHOLESAMI, F. et al. Egonet tensor decomposition for community identification. In: IEEE. *GlobalSIP*. [S.l.], 2016. p. 341–345. Citado 2 vezes nas páginas 16 e 17.
- 2 PROKHORENKOVA, L.; TIKHONOV, A. Community detection through likelihood optimization. In: *The Web Conference*. [S.l.: s.n.], 2019. p. 1498–1508. Citado 2 vezes nas páginas 16 e 37.
- 3 ABBE, E.; BANDEIRA, A. S.; HALL, G. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, IEEE, v. 62, n. 1, p. 471–487, 2015. Citado 6 vezes nas páginas 16, 17, 19, 29, 30 e 37.
- 4 AVRACHENKOV, K. et al. Network partitioning algorithms as cooperative games. *Comp. Social Networks*, v. 5, 2018. Citado 7 vezes nas páginas 16, 17, 18, 21, 33, 35 e 37.
- 5 HAJEK, B. et al. Recovering a hidden community beyond the Kesten–Stigum threshold in $O(|E| \log |V|)$ time. *Journal of Applied Probability*, Cambridge University Press, v. 55, n. 2, p. 325–352, 2018. Citado 2 vezes nas páginas 16 e 17.
- 6 SALES, G. da C. *Majority Vote Community Detection With Dynamic Threshold and Bootstrapped Rounds*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2019. Citado na página 16.
- 7 SWEN, B. *Method for search result clustering*. [S.l.]: Google Patents, 2006. US Patent 11/263,820. Citado na página 16.
- 8 VELUSAMY, S. et al. An efficient ad recommendation system for tv programs. *Multimedia Systems*, Springer, v. 14, n. 2, p. 73–87, 2008. Citado na página 16.
- 9 BEN-DOR, A.; SHAMIR, R.; YAKHINI, Z. Clustering gene expression patterns. *Journal of computational biology*, Mary Ann Liebert, Inc., v. 6, n. 3-4, p. 281–297, 1999. Citado na página 16.
- 10 BLONDEL, V. D. et al. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, IOP Publishing, v. 2008, n. 10, p. P10008, 2008. Citado 4 vezes nas páginas 17, 19, 28 e 35.
- 11 POULIN, V.; THÉBERGE, F. Ensemble clustering for graphs: comparisons and applications. *Applied Network Science*, Springer, v. 4, n. 1, 2019. Citado 6 vezes nas páginas 17, 19, 28, 29, 35 e 36.
- 12 LUXBURG, U. V. A tutorial on spectral clustering. *Statistics and computing*, Springer, v. 17, n. 4, p. 395–416, 2007. Citado 3 vezes nas páginas 17, 28 e 35.
- 13 XU, M. et al. Optimal rates for community estimation in the weighted stochastic block model. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 48, n. 1, p. 183–204, 2020. Citado 2 vezes nas páginas 17 e 36.

- 14 HAJEK, B.; SANKAGIRI, S. Community recovery in a preferential attachment graph. *IEEE Transactions on Information Theory*, IEEE, v. 65, n. 11, p. 6853–6874, 2019. Citado na página 17.
- 15 MAZALOV, V. V. Comparing game-theoretic and maximum likelihood approaches for network partitioning. In: *Trans. Comp. Intelligence*. [S.l.]: Springer, 2018. p. 37–46. Citado 4 vezes nas páginas 17, 21, 33 e 37.
- 16 CHAN, T. F.; JR, P. C.; SZETO, W. On the optimality of the median cut spectral bisection graph partitioning method. *SIAM Journal on Scientific Computing*, SIAM, v. 18, n. 3, p. 943–948, 1997. Citado na página 19.
- 17 GREENE, D.; DOYLE, D.; CUNNINGHAM, P. Tracking the evolution of communities in dynamic social networks. In: IEEE. *Intl. conference on advances in social networks analysis and mining*. [S.l.], 2010. p. 176–183. Citado na página 27.
- 18 GACH, O.; HAO, J.-K. Improving the Louvain algorithm for community detection with modularity maximization. In: SPRINGER. *Artificial Evolution*. [S.l.], 2013. p. 145–156. Citado na página 35.
- 19 YAN, D.; HUANG, L.; JORDAN, M. I. Fast approximate spectral clustering. In: *ACM SIGKDD*. [S.l.: s.n.], 2009. p. 907–916. Citado 2 vezes nas páginas 35 e 36.
- 20 DASKALAKIS, C.; GOLDBERG, P. W.; PAPADIMITRIOU, C. H. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, SIAM, v. 39, n. 1, p. 195–259, 2009. Citado na página 36.
- 21 SAMUELSON, L. *Evolutionary games and equilibrium selection*. [S.l.]: MIT, 1997. Citado na página 36.
- 22 MASSOULIÉ, L. Community detection thresholds and the weak Ramanujan property. In: *ACM symposium on Theory of computing*. [S.l.: s.n.], 2014. p. 694–703. Citado na página 36.

Apêndices

APÊNDICE A – PROVA DO TEOREMA 2.2.1

Demonstração. Considere, mas sem perda de generalização, que um nó j tenha incentivo de migrar da comunidade \mathcal{B} para a comunidade \mathcal{A} . Nosso objetivo é mostrar que o ganho instantâneo derivado de tal movimento é de pelo menos $1/c$. Este resultado, juntamente com o fato de que o potencial é limitado inferior (resp., limitado superior) por $-V^2$ (resp., V^2) implica que os algoritmos convergem em no máximo $2cV^2 = O(V^2)$ passos.

Lembre-se de que n_B é o número de nós na comunidade \mathcal{B} antes de j mover, enquanto que n_A é o número de nós em \mathcal{A} , sendo $n_B + n_A = V$. Lembre-se também de que d_{Bj} e d_{Aj} são os graus de j em \mathcal{B} e em \mathcal{A} , antes e depois do movimento, e \bar{d}_{ij} é o número de nós em i não conectados a j , com $\bar{d}_{Aj} = n_A - d_{Aj}$ e $\bar{d}_{Bj} = n_B - d_{Bj} - 1$.

Seja Δv_{ij} o ganho obtido através da migração do nó j ao sair de i ,

$$\begin{aligned}\Delta v_{Bj} &= ((1 - \alpha)d_{Aj} - \alpha\bar{d}_{Aj}) - ((1 - \alpha)d_{Bj} - \alpha\bar{d}_{Bj}) \\ &= -\alpha(n_A - n_B + 1) + d_{Aj} - d_{Bj}\end{aligned}\tag{A.1}$$

Segue-se de (A.1) que $c\Delta v_{Bj} \in \mathbb{Z}$,

$$c\Delta v_{Bj} = -b(n_A - n_B + 1) + c(d_{Aj} - d_{Bj}).\tag{A.2}$$

Então,

$$\begin{aligned}\Delta v_{Bj} > 0 &\Rightarrow c\Delta v_{Bj} > 0 \Rightarrow \\ c\Delta v_{Bj} &\geq 1 \Rightarrow \Delta v_{Bj} \geq \frac{1}{c}.\end{aligned}\tag{A.3}$$

onde a primeira desigualdade decorre da suposição de que o nó j tem incentivo para migrar (premissa (ii)) e a segunda implicação decorre do fato de que $c\Delta v_{Bj} \in \mathbb{Z}$ (equação (A.2) e premissa (i)).

□

APÊNDICE B – PROVA DO TEOREMA 2.3.1

Demonstração. Consideramos um determinado nó j , que faz parte da comunidade \mathcal{B} . Então, mostramos que o nó j não tem incentivo para migrar de \mathcal{B} para \mathcal{A} para qualquer $\alpha \in [\alpha_0, \alpha_1]$, desde que não tenha incentivo para migrar quando $\alpha = \alpha_0$ e $\alpha = \alpha_1$.

Usamos a mesma terminologia do Teorema 2.2.1 e Apêndice (A.1). De fato, $\Delta v_{Bj}(\alpha)$ é uma função linear de α (veja equação (A.1)), então $\Delta v_{Bj}(\alpha_0) \leq 0$ e $\Delta v_{Bj}(\alpha_1) \leq 0$ implica que $\Delta v_{Bj}(\alpha^*) \leq 0$ para $\alpha^* \in [\alpha_0, \alpha_1]$.

Acabamos de mostrar que o resultado é válido para qualquer nó na comunidade \mathcal{B} . Por simetria, os mesmos argumentos se aplicam aos nós da comunidade \mathcal{A} , o que conclui a prova.

□

APÊNDICE C – PROVA DO TEOREMA 2.3.2

Demonstração. Considere um dado nó j , que faz parte da comunidade \mathcal{B} , e usamos a mesma terminologia do Teorema 2.2.1. Seja B_α e A_α os valores hedônicos do nó j antes e depois de migrar entre comunidades. Segue-se de (2.2) e (A.1) que:

- $\alpha = 0$: equilíbrio se $B_0 \geq A_0$, i.e., $d_{Bj} \geq d_{Aj}$
- $\alpha = 1$: equilíbrio se $B_1 \geq A_1$, i.e.,

$$d_{Bj} - (n_B - 1) \geq d_{Aj} - (n_A + 1 - 1). \quad (\text{C.1})$$

Ao assumir comunidades do mesmo tamanho, $n_A = n_B = n$, a última condição se traduz em $d_{Bj} \geq d_{Aj} - 1$, o que está implícito na condição $d_{Bj} \geq d_{Aj}$ que vale para $\alpha = 0$. Portanto, se duas comunidades têm o mesmo tamanho e compreendem um equilíbrio para $\alpha = 0$, elas também compreendem um equilíbrio para $\alpha = 1$.

□