

Abordagem de redes por *cliques*: aplicação a redes de coautoria

Marcos Grilo Rosa¹, Inácio de Sousa Fadigas¹, Maria Teresinha Tamanini Andrade²,
Hernane Borges de Barros Pereira^{1,3}

¹Universidade Estadual de Feira de Santana, Feira de Santana, Bahia, Brasil

²Instituto Federal de Educação Ciência e Tecnologia, Simões Filho, BA, Brasil

³Programa de Modelagem Computacional, SENAI Cimatec, Salvador, BA, Brasil

{grilo, fadigas}@uefs.br, hbbpereira@gmail.com, tamanini@ifba.edu.br

Abstract. *Some social networks consist of links between groups of actors mutually connected, forming a clique. The classical approach with indexes of Social Network Analysis does not take into account this peculiarity. We introduce new indices of cohesion based approach for cliques, and rethink the concept of network density. We applied this index to two co-authorship networks: one formed by researchers who published in journals of mathematics education and the other by researchers from a post-graduate program in computational modeling. We made a comparative and contextualized analysis to show the applicability and capability indices in the treatment of data from social networks.*

Resumo. *Algumas redes sociais consistem em ligações entre grupos de atores mutuamente conectados que formam uma clique. A abordagem clássica com índices da Análise de Redes Sociais não leva em conta esta particularidade. Introduzimos novos índices de coesão baseados na abordagem por cliques, e ressignificamos o conceito de densidade da rede. Estes foram aplicados a duas redes de coautoria: uma formada por pesquisadores que publicam em periódicos de Educação Matemática e outra por pesquisadores de um Programa de Pós-Graduação em Modelagem Computacional. Uma análise comparativa e contextualizada é feita para mostrar a aplicabilidade e a potencialidade dos índices, no tratamento de dados de redes sociais.*

1. Introdução

A forma como alguns tipos de redes são constituídas, a partir de grupos mutuamente conectados (e. g. grupos dos autores de um artigo), permite a modelagem desses elementos básicos por estruturas da teoria dos grafos chamadas *cliques*, que são grafos (ou subgrafos) cujos vértices são mutuamente conectados.

Redes de coautoria tem sido investigadas por diversos pesquisadores, a exemplo de [Katz and Martin 1997], [Newman 2001] e [Maia and Carenato 2008]. Estas pesquisas usam os índices clássicos da Análise de Redes Sociais, e sobretudo as centralidades para destacar autores na rede. Este artigo pretende ressignificar o conceito clássico de densidade e introduzir novos índices de coesão para redes formadas por *cliques*, que serão aplicados e interpretados em redes de coautoria.

2. Abordagem de redes por *cliques*

Nesta Seção, introduzimos uma abordagem inovadora para tratar redes. A abordagem é baseada na concepção teórica de que as *cliques* isoladas formam as redes por *justaposição* e/ou *sobreposição* de *cliques*. Designamos de justaposição o processo no qual duas *cliques* são ligadas por apenas um vértice comum. Quando a ligação ocorre com dois ou mais vértices comuns, chamamos o processo de sobreposição. A Figura 1(a) apresenta uma *rede de cliques desconectada*, que é a concepção teórica de uma rede formada pelas *cliques* antes do processo de justaposição e/ou sobreposição. A Figura 1(b) exibe um exemplo dos processos de justaposição e sobreposição aplicados à rede da Figura 1(a). As redes resultantes desses processos serão tratadas como *redes de cliques*, e podem ser conectadas ou não.

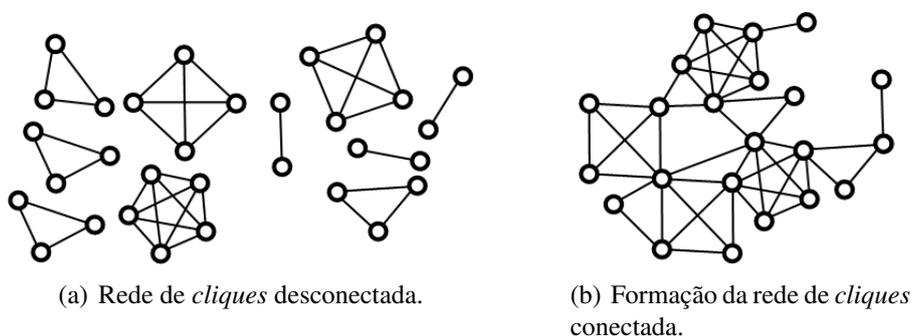


Figura 1. Exemplos de redes de cliques.

3. Índices de coesão baseados na abordagem por *cliques*

A partir da ideia de rede de *clique* desconectada e dos processos de justaposição e/ou sobreposição, que resulta na rede de *clique*, introduzimos alguns índices de coesão: variação de densidade; sobreposição de arestas; fator de redução de vértices e fator de redução de componentes, apenas para exemplificar o caráter inovador da abordagem. A Tabela 1 resume as expressões matemáticas dos índices usados no texto.

Tabela 1. Definição dos índices usados no texto.

Índice	Δ	Δ_{q0}	$v(\Delta)$	S_p	FRV	FRC
Definição	$\frac{2 \mathcal{E} }{n(n-1)}$	$\frac{\sum_{i=1}^{n_q} q_i(q_i-1)}{n_0(n_0-1)}$	$\frac{\Delta - \Delta_{q0}}{\Delta_{q0}}$	$\frac{ \mathcal{E}_0 - \mathcal{E} }{ \mathcal{E}_0 }$	$\frac{n_0 - n}{n}$	$\frac{n_q - C_c}{n_q - 1}$

3.1. Variação de densidade

A expressão matemática para o cálculo da densidade Δ para redes não dirigidas é dada na Tabela 1 que relaciona a quantidade de arestas existentes na rede ($|\mathcal{E}|$) ao número máximo possível destas, dado por $(n(n-1)/2)$, onde n é o número de vértices da rede.

Consideremos uma rede de *cliques* na qual o tamanho (isto é, a quantidade de vértices) de pelo menos uma *clique* é maior do que ou igual a 3. Quando as *cliques* são desconectadas (Figura 1(a)), o número de arestas é dado pela soma da quantidade de arestas em cada *clique*. Chamando de n_q o número de *cliques*, q_i o tamanho da i -ésima

clique e n_0 o número total de vértices da rede de *cliques* desconectada, a densidade Δ_{q_0} expressa pela equação da Tabela 1. Portanto, para uma rede de *cliques* desconectada a densidade $\Delta_{q_0} \neq 0$, uma vez que o número de arestas $|\mathcal{E}| \neq 0$.

Podemos também comparar a densidade da rede de *cliques* com a densidade da rede de *cliques* desconectada (Figura 1(a)). A variação da densidade da rede de *cliques* $v(\Delta)$ é dada na Tabela 1. A expressão captura a variação da densidade relativa a Δ_{q_0} , ou seja, mede o “adensamento” da rede em relação a seu estado inicial (rede de *cliques* desconectada).

3.2. Sobreposição de arestas e fator de redução de vértices

A taxa de sobreposição S_p dada na Tabela 1 compara a quantidade de arestas da rede de *cliques* desconectada com a quantidade de arestas após o processo de justaposição e/ou sobreposição. Esta taxa é definida pela $S_p = \frac{A}{n}$ e é aplicável tanto a redes conectadas quanto a redes desconectadas. O fator de redução de vértices (FRV), cuja expressão está na Tabela 1, pode ser usado para comparar a quantidade de vértices da rede de *cliques* desconectada, com a quantidade de vértices da rede resultante da justaposição e/ou sobreposição.

3.3. Fator de redução de componentes

O fator de redução de componentes (FRC) mede o quanto os componentes são conectados para formar componentes (C_p) maiores e reduzir o número destes. A normalização é feita tomando como base o número mínimo de componentes que pode resultar, que é 1. Como n_q designa o número de componentes da rede de *cliques* desconectada, se C_c é o número de componentes da rede final, então, o fator pode ser dado pela expressão da Tabela 1

4. Aplicação em redes de coautoria

Uma vez descritos os índices, empregando uma nova abordagem, escolhemos duas redes de coautoria para os cálculos e interpretações. Uma rede é formada pelos autores que publicaram em seis periódicos de divulgação da Matemática (Rede EM). A outra rede é formada por pesquisadores de um programa de pós-graduação em modelagem computacional (Rede PPG). As redes de coautoria foram construídas de forma que cada *clique* de coautores está conectada a outra se pelo menos tiver um autor (A) em comum. Para efeito de interpretação da coautoria como uma forma de colaboração, excluímos das redes os autores que só publicaram isoladamente. A Tabela 2 resume as principais quantidades das redes.

Tabela 2. Quantitativos e índices de coesão para as Redes EM e PPG

	Quantitativos					Índices de coesão (abordagem por cliques)				
	n	A	A/n	C_p	$MaxC_p(\%)$	Δ	$v(\Delta)$	S_p	FRV	FRC
Rede EM	588	1.751	2,98	160	7,14	0,004	7,878	0,093	0,664	0,874
Rede PPG	792	3.234	4,08	8	90,28	0,007	15,444	0,500	0,755	0,994

5. Resultados e interpretações

Os índices de coesão calculados para as duas redes estão na Tabela 2. A variação de densidade mede o “adensamento” da rede, ou seja, o quanto os grupos de coautores se aglutinaram para formar a rede. O índice mede simultaneamente a variação de vértices e

a variação de arestas, frente à rede desconectada, mas diferente da densidade clássica, a sua interpretação é relativa a “grupos” e suas ligações, e não apenas a vértices e arestas. Os resultados da Tabela 2 mostram que a Rede PPG exibe $v(\Delta)$ maior que aquele para a Rede EM. O valor teórico máximo para a variação de densidade ocorre quando a rede torna-se uma única *clique*. Para a Rede EM, este valor máximo é da ordem de 2.200, portanto, o valor encontrado (7,8784) corresponde a apenas 0,4% do máximo. No caso da Rede PPG, a variação máxima da densidade chega a 1.223, e o valor encontrado (15,4440) corresponde a 1,3% do máximo. Comparativamente, podemos dizer que a Rede PPG é mais de três vezes mais “adensada” que a Rede EM. Se compararmos apenas a densidade, este fator é pouco maior que 2 vezes.

A sobreposição de arestas corresponde ao peso das arestas na análise clássica de redes. Na abordagem por *cliques*, reflete o percentual de grupos, isolados inicialmente, que se unem para formar a rede. Este índice, indica que na Rede PPG há, além de uma maior produção científica por pares de autores, uma maior integração entre os grupos de coautores, em relação à Rede EM. O fator de redução de vértices, por sua vez, está diretamente ligado à relação autor/vértice da Tabela 2. O índice indica o percentual de autores com mais de uma publicação; por outro lado, na abordagem por cliques, indica também a ligação entre os grupos via vértices (justaposição). Como o percentual de sobreposição de arestas é muito baixo para a rede EM, significa que nesta rede as ligações entre os grupos é feita prioritariamente por justaposição, o contrário do que ocorre com a rede PPG.

Quanto ao fator de redução de componentes, os valores obtidos para as duas redes não são discrepantes. Mesmo que a Rede EM tenha exibido 160 componentes contra apenas 8 da Rede PPG, a diferença percentual é em torno de 12%. Por outro lado, quase todos os grupos de coautoria tiveram pelo menos um vértice em comum (de ligação) na Rede PPG (99,4%), enquanto que na Rede EM este valor foi de 87,4%.

6. Considerações finais

A abordagem por *cliques* em redes de coautoria permitiu uma forma de tratamento de dados sociais mais adequado à estrutura topológica da rede. A análise das redes, a partir de poucos índices de coesão, já permite a emergência de novas interpretações. A exemplo, o índice que mede a sobreposição de arestas, juntamente com o fator de redução de vértices, permitem clarificar a forma como ocorre o processo de justaposição e sobreposição para a formação da rede.

A pesquisa inicial com os índices de coesão mostrou que potencialmente outros índices podem ser acrescentados, bem como investigada a dinâmica de crescimento das redes. Outro aspecto que destacamos é a aplicabilidade da abordagem por *cliques* em outras redes sociais que tenham a mesma estrutura, a exemplo da rede de ator-filme.

Referências

- Katz, J. S. and Martin, B. R. (1997). What is research collaboration? *Research Policy*.
- Maia, M. F. S. and Caretato, S. L. (2008). Co-autoria como indicador de colaboração científica. *Perspectivas em Ciência da Informação*, 13(12):18–31.
- Newman, M. E. J. (2001). The structure of scientific collaboration networks. *PNAS*, 98(2):404–409.