

REDES DE INTERAÇÃO CIENTÍFICA: identificação de comunidades de autores na UFRJ

Roberto Mario Lovón Canchumani

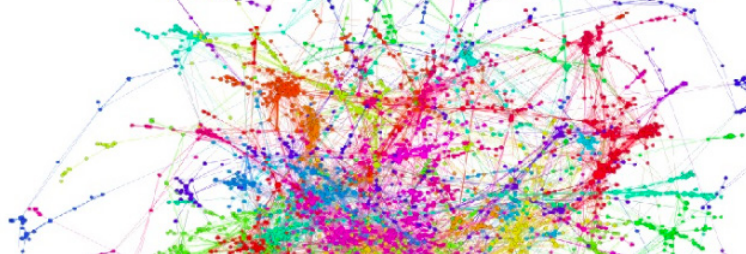
Instituto Brasileiro de Informação em Ciência e Tecnologia
Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil
rlovon@gmail.com

1 INTRODUÇÃO

O desenvolvimento de pesquisas no campo de redes complexas tem apresentado, nas últimas décadas, significativos avanços na descrição e análise dos traços topológicos e dinâmicos de redes de tipo social, biológica ou tecnológica (BARABÁSI et al., 2002). Neste cenário, um dos aspectos que tem concentrado particular atenção tem sido a existência de subconjuntos de nós fortemente entrelaçados, às vezes conectados e outras vezes desconectados da rede. Neste último caso, nos deparamos com agrupamentos ou comunidades (*clusters*) que se encontram ligadas ou não à rede (FORTUNATO, 2010).

Essas comunidades cumprem papel relevante nas propriedades de estruturas complexas. Identificar e analisar a natureza das comunidades aí ocultas é uma importante tarefa para revelar a organização informal e a natureza dos fluxos de informação que acontecem dentro desses sistemas complexos (PALLA et al., 2005). É importante destacar que tais comunidades são compostas por diferentes subconjuntos de nós, os quais se caracterizam por estarem densamente ligados mais entre si do que em relação ao resto da rede. A identificação dessas comunidades facilita a compreensão das redes.

Diversas abordagens foram desenvolvidas para identificar comunidades, nas quais se adotam funções específicas para descrever características dos grupos a serem determinados (CANCHUMANI et al.,



2017; BARBOSA et al., 2011). Este trabalho apresenta os resultados de uma proposta de identificação de comunidades de autores realizada na Universidade Federal do Rio de Janeiro (UFRJ). O objetivo do estudo foi identificar a tendência de agrupamento de autores da instituição ao longo do tempo, assim como também observar características mais específicas das principais comunidades de autores identificadas.

2 METODOLOGIA

Foi realizado um levantamento da produção bibliográfica, envolvendo artigos publicados em periódicos no período de 2001 a 2012, na base de dados institucional SIGMA.UFRJ. A busca na base considerou os seguintes critérios: Produção Intelectual => artigos completos publicados em periódicos => autores. A partir desses critérios foi possível extrair 44.233 registros da produção da UFRJ de artigos publicados em periódicos no período de 2001 a 2012. Os registros com suas informações foram exportados para matrizes multidimensionais, um processo que usou ferramentas computacionais, como software para cruzamento de dados (*Excel*) e geração/visualização de redes/mapas (*Gephi*). A interação entre autores foi analisada a partir das informações de coautoria e as métricas de grau e modularidade/transitividade.

3 RESULTADOS

A população identificada na produção acadêmica no período de 2001 a 2012 é constituída por 28.248 autores com vínculo formal com a UFRJ, responsáveis pelos 44.233 artigos publicados no período. Nas redes de interação de coautoria, em média, cada autor da UFRJ mantém colaboração com 6,73 pesquisadores.

Um olhar por triênios mostrou o seguinte: no triênio 2001-2003, o grau médio foi de 4,33, ou seja, cada autor colaborou, em média, com pouco mais de quatro autores, e a transitividade, isto é, a tendência que os autores têm de se agruparem (coeficiente de *clustering*) foi de 0,76. Em 2004-2006, o grau médio foi de 4,87 e o coeficiente de *clustering* foi de 0,77. No período seguinte, 2007-2009, o grau médio foi de 5,12 e a



transitividade foi de 0,78. Já o último triênio apresentou grau médio de 5,57 e transitividade de 0,79. Essas medidas indicam que, ao longo dos períodos, novas colaborações foram estabelecidas formando novos agrupamentos/comunidades. Em outras palavras, ao longo dos triênios, a tendência de todos os nós (autores) se agruparem aumenta.

A fim de verificar a rede de autores com participação mais intensa em termos de colaboração na instituição ao longo do tempo, identificamos a formação de diferentes subgrupos de pesquisadores que formam comunidades (*clusters*) nos triênios 2001-2003 e 2009-2012, focalizando nossa atenção na maior comunidade de cada um desses períodos.

3.1 COMUNIDADES DE AUTORES NO TRIÊNIO 2001-2003

A rede de coautoria da UFRJ, no período 2001-2003, é composta por 5.695 autores. A identificação de agrupamentos dessa população, aplicando o algoritmo Blondel et al. (2008), encontrou um total de 40 comunidades, envolvendo 3.736 nós (autores), identificadas na Figura 1-A através das cores. Os nós estabeleceram entre si 11.304 relações (colaborações), representadas na rede por arestas. O grau médio foi de 6,05, ou seja, em média, estes autores participaram de publicações com pouco mais de seis autores e a transitividade (coeficiente de *clustering*) de 0,76.

A maior comunidade de autores identificada no período de 2001-2003 é constituída por 236 autores (o que representa 7,04% do total de autores das 40 comunidades) e, 689 relacionamentos, arestas (Figura 1-B). O grau médio é de 5,24, ou seja, cada autor colabora, em média, com mais de cinco pesquisadores. O coeficiente de transitividade apresentado é de 0,79.

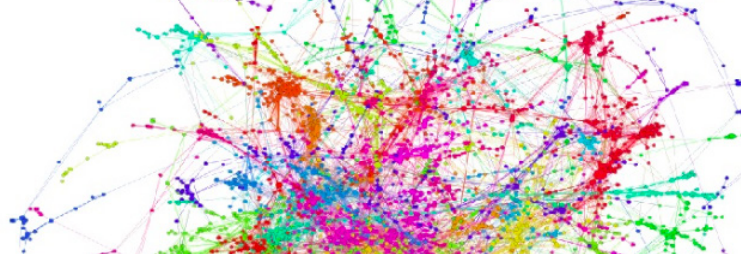
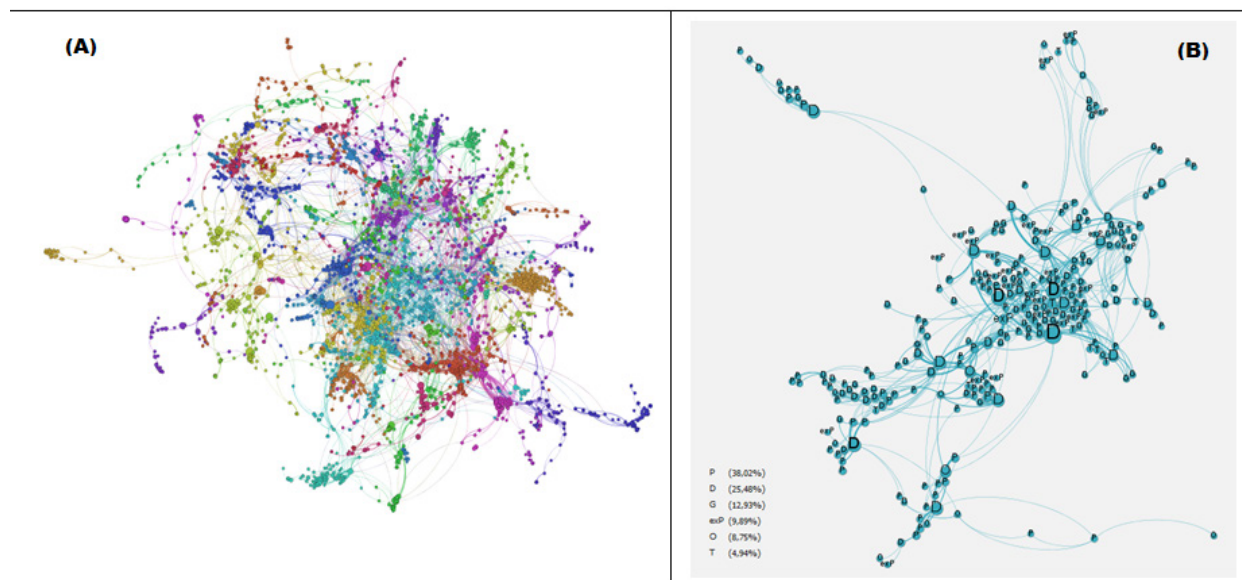


FIGURA 1 - (A) GRAFO DAS COMUNIDADES DE AUTORES DA UFRJ DE PUBLICAÇÕES EM PERIÓDICOS IDENTIFICADAS NO PERÍODO 2001-2003. (B) GRAFO DA MAIOR COMUNIDADE DE AUTORES DE ARTIGOS EM PERIÓDICOS DA UFRJ NO PERÍODO 2001-2003

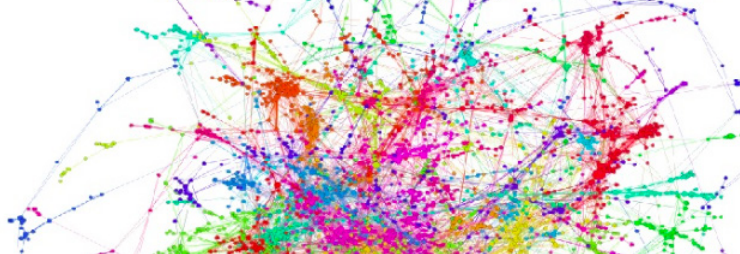


Fonte: Elaborado pelo autor, 2018.

A classificação de autores dessa comunidade por tipo de vínculo mostra que a maioria é composta por alunos de pós-graduação - P (38,02%), seguida por docentes - D (25,48%) e alunos de graduação - G (12,93%). Alunos egressos da pós-graduação (exP) representam 9,89% e os autores com outros tipos de vínculo (O) 8,75%, podendo ser professor egresso, técnico egresso ou egresso da graduação. Já os técnicos (T) representam 4,94%.

Do total de autores dessa comunidade, identificamos os 20 que apresentam as maiores medidas de centralidade de grau, ou seja, são autores que mais interagem, tem maior número de colaborações (Tabela 1). Verifica-se que dos 20 autores identificados, 16 são docentes. Os outros quatro autores têm os seguintes tipos de vínculo: egresso da pós-graduação (exP), aluno de pós-graduação (P), técnico (T) e um que possui outros tipos de vínculo (O).

No que se refere às áreas dos autores, verificamos que 10 deles atuam na grande área de Ciências Biológicas, dois nas Ciências da Saúde e um nas Ciências Humanas. Os outros autores identificados atuam em duas áreas do conhecimento, assim distribuídos: quatro deles em Ciências



Biológicas e Ciências da Saúde; dois em Ciências Agrárias e Ciências da Saúde e; um autor em Ciências Biológicas e Ciências Humanas.

TABELA 1 - VÍNCULO E GRANDE ÁREA DOS AUTORES COM MAIORES MEDIDAS DE CENTRALIDADE DE GRAU DA MAIOR COMUNIDADE IDENTIFICADA DAS PUBLICAÇÕES DA UFRJ EM 2001-2003

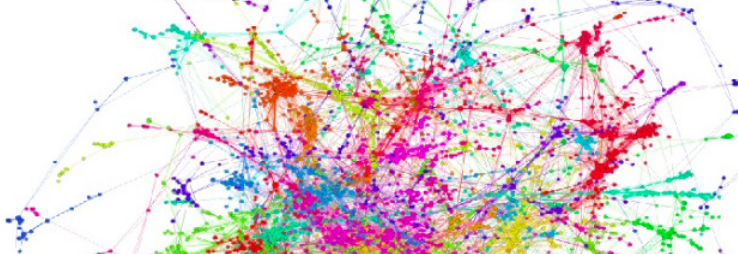
Id	Vínculo*	Grau	Áreas de Atuação
330	D	46	Ciências Biológicas, Ciências da Saúde
2139	D	34	Ciências Biológicas, Ciências da Saúde
342	D	26	Ciências Biológicas
339	exP	24	Ciências Biológicas, Ciências da Saúde
343	O	20	Ciências Biológicas
344	D	19	Ciências Biológicas
2140	D	19	Ciências da Saúde
268	D	17	Ciências Biológicas
406	D	16	Ciências Biológicas
2042	D	15	Ciências Agrárias, Ciências da Saúde
2320	D	15	Ciências da Saúde, Ciências Biológicas
445	D	14	Ciências Biológicas
540	D	14	Ciências Biológicas
2156	T	14	Ciências da Saúde
223	D	13	Ciências Biológicas
412	D	12	Ciências Biológicas, Ciências Humanas
1924	D	12	Ciências da Saúde, Ciências Agrárias
338	P	11	Ciências Humanas
341	D	11	Ciências Biológicas
541	D	11	Ciências Biológicas

Fonte: Elaborado pelo autor, 2018.

* Tipos de Vínculo: D (Docente); P (Discente de Pós-graduação); exP (Egresso Discente de Pós-graduação); T (Técnico); O (Outros)



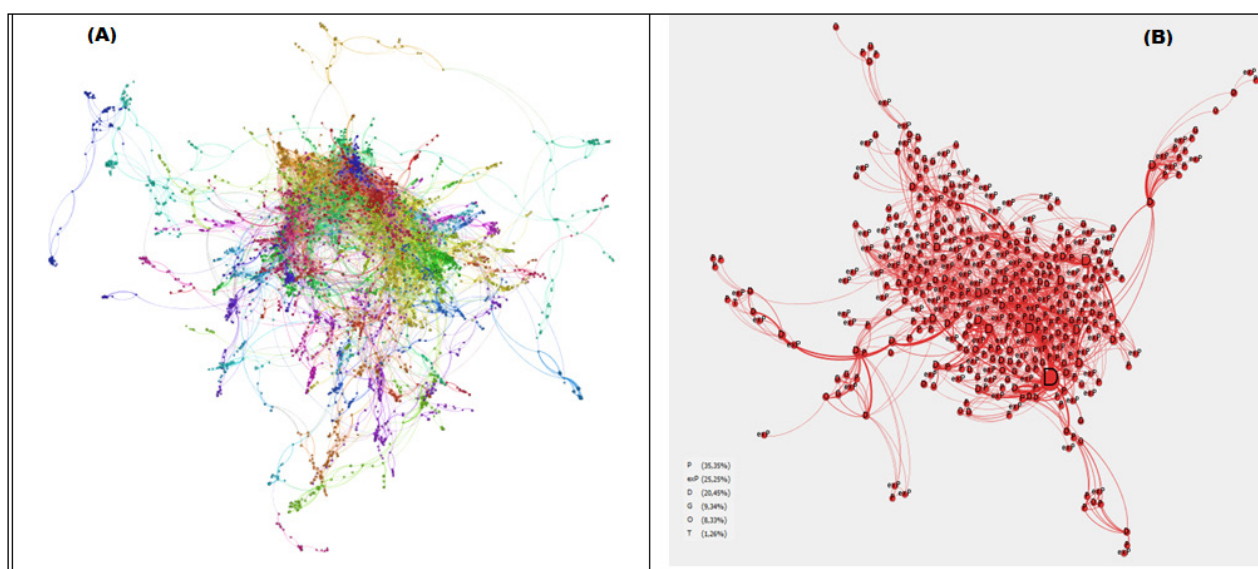
Uma visão detalhada da rede em torno do grupo de autores mais centrais revela algumas informações importantes (Figura 2). O tamanho do nó (e do Id) é proporcional ao número de colaborações (grau do nó) e cada tipo de autor é representado por uma determinada cor. Podemos observar claramente que os docentes mais centrais (id= 330 e 2139) concentram em seu entorno diferentes tipos de autores, formando o principal núcleo de colaboração dessa comunidade. Outros docentes, por sua vez, concentram sua colaboração com algum tipo de autores. Por exemplo, os docentes id= 540 e 2042 colaboram, basicamente, com alunos de pós-graduação.



3.2 COMUNIDADES DE AUTORES NO TRIÊNIO 2009-2012

A rede de autores da UFRJ, no período 2010-2012, é composta por 7.503 autores (total de autores em artigos com colaboração). A identificação de agrupamentos dessa população, aplicando o algoritmo Blondel et al. (2008), apontou um total de 40 comunidades, envolvendo 5.169 autores, os quais mantiveram entre si 19.328 conexões, representadas na rede por arestas (Figura 3-A). O grau médio foi de 7,47 e a transitividade (coeficiente de *clustering*) de 0,78.

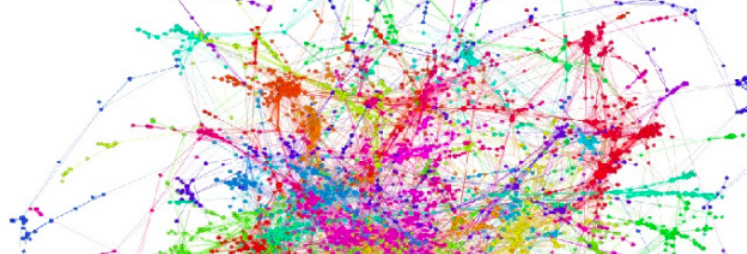
FIGURA 3 - (A) GRAFO QUE MOSTRA AS 40 COMUNIDADES DE AUTORES DA UFRJ IDENTIFICADAS NO PERÍODO 2010-2012. (B) GRAFO QUE DESTACA A MAIOR COMUNIDADE DE AUTORES DA UFRJ IDENTIFICADA NO PERÍODO 2010-2012



Fonte: Elaborado pelo autor, 2018.

Note-se que o número de comunidades é igual ao identificado no período 2001-2003, entretanto, o número de autores e conexões aumentou significativamente. Isto pode indicar que o aumento da colaboração ocorreu pelo incremento dos agrupamentos já existentes em períodos anteriores.

A maior comunidade de autores identificada no período de 2010-2012 é constituída por 396 autores (o que representa 7,66% do total de autores das 40 comunidades), 160 autores a mais que no triênio 2001-2003. Estes autores estabeleceram 1.607 relacionamentos, arestas (Figura



3-B). O grau médio é 8,12, isto é, cada autor colabora, em média, com pouco mais de oito pesquisadores, e o coeficiente de *clustering* é de 0,81.

Ao verificar os autores dessa comunidade de acordo com seus respectivos tipos de vínculo, vemos que os grupos mais representativos são os alunos de pós-graduação (35,35%), seguido pelos egressos da pós-graduação (25,25%) e docentes (20,45%). Os estudantes de graduação representam 9,34%, os autores com outros tipos de vínculo 8,33% e os técnicos 1,26%. Desse total, identificamos os 20 autores que apresentam as maiores medidas de centralidade (Tabela 2).

TABELA 2 - VÍNCULO E GRANDE ÁREA DOS AUTORES COM MAIORES MEDIDAS DE CENTRALIDADE DE GRAU DA MAIOR COMUNIDADE IDENTIFICADA DAS PUBLICAÇÕES DA UFRJ EM 2010-2012

Id	Vínculo*	Grau	Áreas de Atuação
711	D	80	Ciências Biológicas
536	D	47	Ciências Biológicas
2712	D	44	Ciências Biológicas
180	D	43	Ciências Biológicas
179	D	37	Ciências Biológicas
678	D	37	Ciências Biológicas
350	D	36	Ciências Biológicas
9049	D	36	Ciências Biológicas
167	D	34	Ciências Biológicas
168	D	34	Ciências Biológicas
334	D	33	Ciências Biológicas
177	P	32	Ciências Biológicas
1436	D	27	Ciências Biológicas
1623	D	27	Ciências Biológicas
563	D	25	Ciências Biológicas
564	D	25	Ciências Biológicas
1398	D	25	Ciências Biológicas, Ciências Exatas e da Terra
1714	D	25	Ciências Biológicas, Ciências Humanas
1608	D	23	Ciências Biológicas
1609	D	23	Ciências Biológicas, Ciências da Saúde

Fonte: Elaborado pelo autor, 2018.

* Tipos de Vínculo: D (Docente); P (Discente de Pós-graduação)

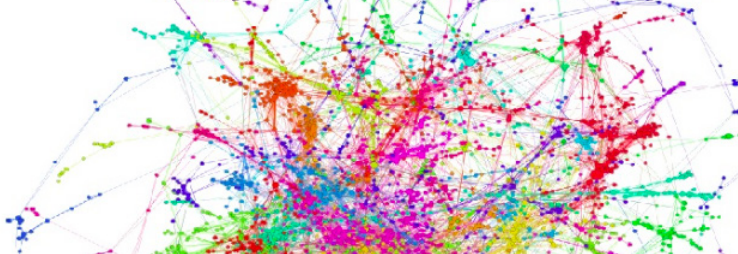
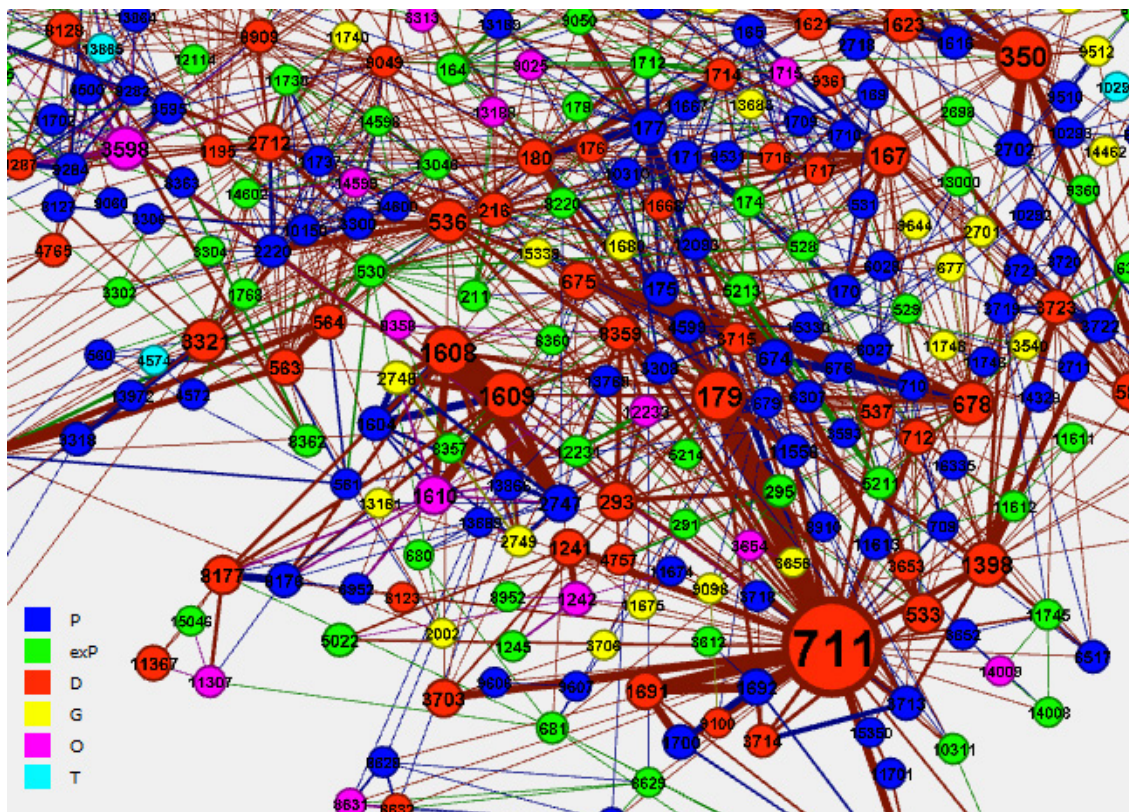


FIGURA 4:- GRAFO QUE DESTACA OS AUTORES MAIS CENTRAIS DA MAIOR COMUNIDADE DA UFRJ IDENTIFICADA EM 2010-2012



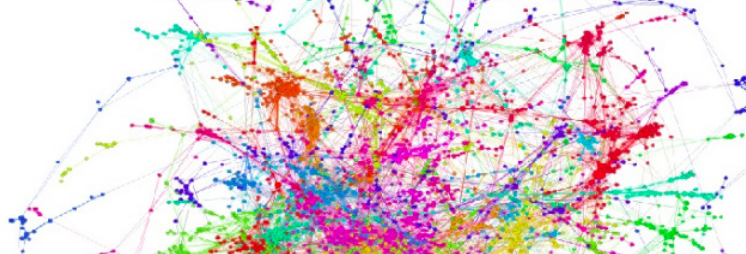
Fonte: Elaborado pelo autor, 2018.

Verifica-se que 19 deles são docentes e um é aluno de pós-graduação. Em relação às áreas dos autores, constatamos que todos atuam na área de Ciências Biológicas, sendo que três deles atuam também nas seguintes áreas: Ciências da Saúde; Ciências Exatas e da Terra e; Ciências Humanas.

Uma visão detalhada da rede em torno do grupo de autores mais centrais revela algumas informações interessantes (Figura 4). É possível notar forte interação entre os docentes identificados com os números 711 e 179. O primeiro mantendo, em seu entorno, colaborações com outros docentes e, o último, colaborando mais intensamente com alunos de pós-graduação.

4 CONSIDERAÇÕES FINAIS

A maior parte das redes do mundo real se caracteriza por conter comunidades muito entrelaçadas e ocultas, nas quais seus membros



formam subgrupos, configurando uma imbricada rede entre todos eles. A identificação de comunidades possibilita definir e melhor conhecer os subgrupos em função da estrutura de escolha de seus enlaces, isto é, com base na premissa de que os membros de cada subgrupo tendem a escolher os mesmos colaboradores e serem escolhidos pelos mesmos autores. Dessa forma, a permanência em um subgrupo estabelece-se em função das similaridades de escolha, dadas e recebidas, por cada autor.

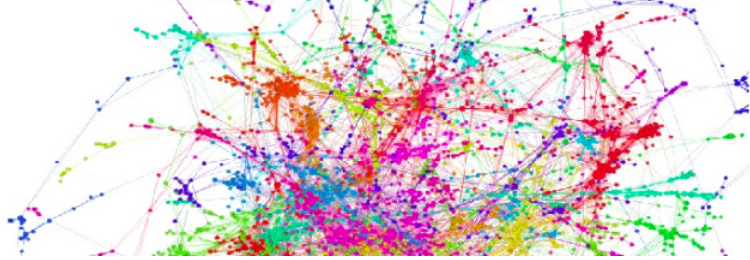
Nesse sentido, as métricas de grau e de transitividade mostraram que, ao longo do período analisado, novas colaborações entre autores da UFRJ foram estabelecidas configurando novos agrupamentos, tornando, dessa maneira, maior a transitividade da rede. Em outras palavras, ao longo dos triênios, a tendência dos autores da UFRJ se agruparem aumenta. Conhecer mais detalhadamente como essa interação acontece, e o papel que desempenha no fazer científico da instituição, é importante para analisar o grau de eficiência/participação dos recursos humanos envolvidos.

O estudo das comunidades de autores pode ser utilizado na gestão institucional, por exemplo, para determinar os principais pesquisadores, número de publicações por grupo, produtividade em relação ao tamanho do grupo, visibilidade e colaboração com outros grupos, assim como também a dinâmica de interação dos grupos de autores ao longo do tempo. Considerações acerca das especificidades das diferentes áreas do conhecimento envolvendo comunidades de autores é aspecto a ser abordado em futuros trabalhos.

REFERÊNCIAS

BARABÁSI, A.; JEONG, H; NÉDA, Z.; RAVASZ, E; SCHUBERT, A; VICSEK, T. Evolution of the social network of scientific collaborations. **Physica A**, v. 311, p. 590-614, 2002.

BARBOSA, D.; AVELINO, L; SOUZA, R.; OLIVEIRA, C.; JUSTEL, C. Medidas de centralidade e detecção de comunidades em rede de co-autoria. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 43., 2011, Ubatuba. **Anais... SOBRAPO**: Ubatuba, 2011. p. 1-10.



BLONDEL, V.; GUILLAUME, J.; LAMBIOTTE, R; LEFEBVRE, E. Fast unfolding of communities in large networks. **J. Stat. Mech**, p.10008, 2008.

CANCHUMANI, R. M. L.; LETA, J.; FIGUEIREDO, A. M.D. Domínios Científicos: mapeamento de áreas do conhecimento da Universidade Federal do Rio de Janeiro. **Inf. & Soc: Est.**, v.27, n. 2, p.199-218, 2017.

FORTUNATO, S. Community detection in graphs. **Physics Reports**, v. 486, n. 3, p. 75-174, 2010.

PALLA, G.; DERÉNYI, I.; FARKAS, I.; VICSEK, T. Uncovering the overlapping community structure of complex networks in nature and society. **Nature**, v. 435, p. 814-818, 2005.