

Uma Visão da Ciência das Redes sobre o Instituto Nacional de Ciência e Tecnologia em Informação Quântica (INCT-IQ)

A network science viewpoint on the National Institute of Science and Technology of Quantum Information (INCT-IQ)

Rute Oliveira^{*1}, Liliana Sanz², Rafael Chaves^{1,3}

¹Universidade Federal do Rio Grande do Norte, Instituto Internacional de Física, Natal, RN, Brasil.

²Universidade Federal de Uberlândia, Instituto de Física, Uberlândia, MG, Brasil.

³Universidade Federal do Rio Grande do Norte, Faculdade de Ciências e Tecnologias, Natal, RN, Brasil.

Recebido em 1 de julho de 2022. Revisado em 5 de agosto de 2022. Aceito em 9 de agosto de 2022.

A ciência das redes é um campo multidisciplinar que oferece um arcabouço amplo para se estudar propriedades estatísticas de uma variedade de fenômenos. No cerne do seu sucesso, está o fato de que os sistemas, por mais complexos que sejam seus constituintes ou interações, podem ser representados por um simples grafo, um conjunto de nós conectados por arestas. Nesta abordagem, processos de natureza muito diferentes, como a internet, colaborações científicas, ou redes de proteínas, se tornam semelhantes do ponto de vista da rede, o que nos permite não somente entender de maneira unificada as redes naturais mas também otimizar e projetar redes artificiais mais eficientes. Dentro deste contexto, este artigo tem dois objetivos. Primeiramente, apresentar os principais conceitos da ciência das redes, tais como grafos, propriedade de mundo pequeno, distribuição de conectividade entre outros, assim como alguns dos principais modelos de redes já propostos. O segundo objetivo é aplicar este ferramental para analisar uma rede real, mais precisamente a rede de pesquisadores do Instituto Nacional de Ciência e Tecnologia de Informação Quântica. Nossos resultados mostram que do ponto de vista estatístico a rede estudada é bem descrita por uma lei de potência truncada, com um alto grau de interconectividade entre os participantes. Um aglomerado, formado por 8 comunidades menores, contém 85% dos cientistas da rede. O número médio de colaborações da rede é próximo de 5 e a média de artigos publicados está acima de 13 durante o período de duração do projeto. A rede possui um alto grau de agregação, com valor de $\langle C \rangle = 0.4$, mostrando que os colaboradores de um dado cientista também tendem a colaborar entre si.

Palavras-chave: Redes Complexas, Ciência das Redes, Informação Quântica.

Network science is a multidisciplinary field that offers a broad framework for studying statistical properties of a variety of phenomena. At the core of its success is the fact that systems, in spite of the complexity of their constituents or interactions, can be represented as a simple graph, a set of nodes connected by edges. In this approach, processes of a very different nature, such as the internet, scientific collaborations, or protein networks, become similar from a network point of view, which allows not only to understand natural networks in a unified way, but also to optimize and design more efficient artificial networks. Within this context, this article has two objectives. First, present the main concepts of network science, such as graphs, the small-world property, connectivity distribution, among others, as well as some of the main network models proposed in the literature. The second objective is to apply network science to analyze a real network, more precisely the network of researchers from the National Institute of Science and Technology of Quantum Information. Our results show that from a statistical point of view the studied network is well described by a truncated power law, with a high degree of interconnectivity among the participants. A cluster, formed by 8 smaller communities, contains 85% of the network's scientists. The network's average number of collaborations is close to 5 and the average of articles published is above 13 in the analyzed period of time. The network has a high clustering coefficient, with a value of $\langle C \rangle = 0.4$, showing that the collaborators of a given scientist also tend to collaborate with each other.

Keywords: Complex Networks, Network Science, Quantum Information.

1. Introdução

A ciência das redes é um campo multidisciplinar que oferece um arcabouço teórico em comum – tanto de

linguagem, conceitos e ferramentas – para se analisar as propriedades estatísticas dos mais variados sistemas, abrangendo desde as ciências sociais [1], biológicas [2, 3] e físicas [4–6] até mesmo redes econômicas [7–9] e ecológicas [10], dentre outras [11, 12]. No cerne deste sucesso está o fato de que diferentes sistemas podem ser

* Endereço de correspondência: ooliveirarm@gmail.com

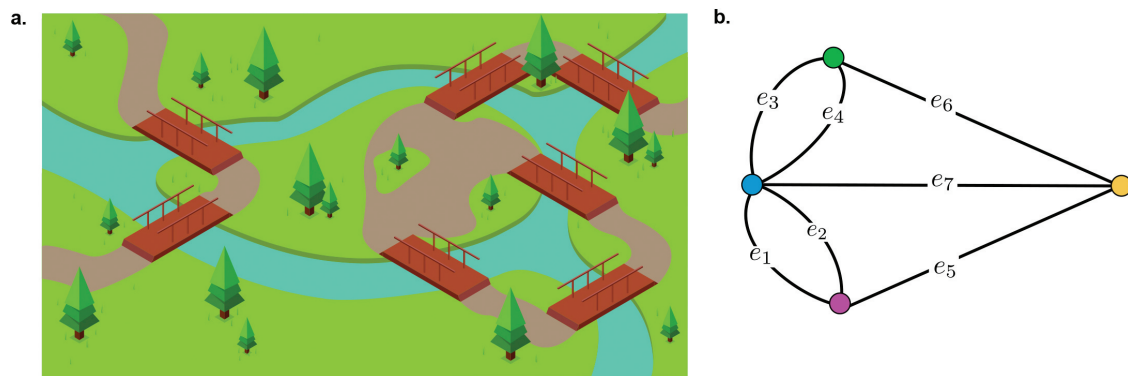


Figura 1: a. Representação artística do problema das 7 pontes de Königsberg. b. Representação utilizando grafos do mesmo problema. As esferas são os vértices que correspondem as ilhas e as ligações entre elas são as arestas que representam as pontes de Königsberg. Portanto, o grafo possui 4 vértices e 7 arestas $E = [e_1, e_2, e_3, e_4, e_5, e_6, e_7]$.

representados através de grafos [13], isto é, um conjunto de nós/vértices conectados por meio de arestas. Do ponto de vista da ciência de redes, são justamente estas conexões que importam, independentemente das várias particularidades de cada sistema ou do comportamento individual de cada um dos seus constituintes. Fenômenos díspares, tais como a internet, redes de colaborações científicas, a malha aeroportuária ou redes de proteínas, podem, do ponto de vista de redes, se comportar de maneira muito similar.

Historicamente, o início da ciência das redes pode ser traçado ao ano de 1735, quando o matemático suíço Leonard Euler resolveu o famoso problema das sete pontes de Königsberg (atualmente Kaliningrado) por meio da teoria de grafos. A Figura 1(a) mostra uma representação artística da antiga Königsberg, na qual vemos o rio Pregel e as ilhas por ele formadas, que contavam, com sete pontes que as interconectavam ao restante da cidade. Um famoso quebra cabeças da época consistia em responder se seria possível atravessar todas as ilhas, passando-se somente um vez por cada uma das pontes, e ao final retornar ao ponto de origem. O problema se manteve sem resposta até que Euler mapeou o problema em um simples grafo, no qual as porções de terra eram os vértices e as pontes representavam arestas, como ilustra a Fig. 1(b). Caso somente um dos vértices apresentasse um número ímpar de arestas, a caminhada imaginada pelos habitantes da cidade seria possível. Mas como todas as ilhas tinham um número ímpar de conexões, Euler acabou por responder em negativo à famosa pergunta [14].

Seria, no entanto, apenas mais de dois séculos após, que a teoria de redes e grafos veria um avanço significativo, quando os matemáticos húngaros Paul Erdős e Alfréd Rényi apresentaram um novo conceito de grafos aleatórios, no qual quaisquer dois nós de um grafo tem a mesma probabilidade de se conectarem [15]. Apesar desse caráter completamente aleatório, a rede probabilística de Erdős-Rényi dava origem a uma propriedade muito interessante: o menor caminho médio

entre quaisquer dois nodos era proporcional ao logaritmo do número total de nodos da rede. Essa propriedade, chamada de mundo pequeno, já havia sido proposta na década de 1920, em um conto do escritor húngaro Frigyes Karinthy [16]. Em suas próprias palavras:

Um jogo fascinante nasceu da discussão. Um deles sugeriu realizar o seguinte experimento para provar que a população da Terra é muito mais próxima hoje do que jamais fora. Deveríamos selecionar qualquer pessoa dos 1,5 bilhão de habitantes da Terra – qualquer um, de realmente qualquer lugar. Ele aposta que, usando não mais que cinco indivíduos, um dos quais seja conhecido pessoal dele, poderia contatar o indivíduo selecionado recorrendo a nada além de sua rede de pessoas conhecidas.

Conhecida como a teoria dos “seis graus de separação”, já que implicava que bastariam seis laços de amizade para que duas pessoas se conectarem, tal proposta viria a ser empiricamente comprovada na década de 60, quando a teoria de grafos passou a ser utilizada para modelar interações sociais. Em particular, o psicólogo social Stanley Milgram publicou seu famoso artigo [17], no qual executou um elaborado experimento de envio de cartas entre desconhecidos e que lhe permitiu provar o caráter de mundo pequeno desta rede social.

Ao longo dos anos seguintes, a ciência de redes e a teoria de grafos passaram a ser utilizadas para estudar os mais variados fenômenos físicos e biológicos (ver, por exemplo, [11, 12]). E apesar do sucesso do modelo de Erdős-Rényi em reproduzir a propriedade de mundo pequeno, notou-se que tal modelo falhava em reproduzir várias outras características que passaram a ser observadas em diversas redes reais. Em particular, notou-se que o chamado coeficiente de agregação, o qual mede as conexões entre os vizinhos de um dado nó da rede, é tipicamente grande nas mais variadas redes. Em uma rede de amigos, por exemplo, seria de se

esperar que todos se conhecessem entre si. Ao contrário, a rede aleatória de Erdős-Rényi previa que a chance de que dois de seus amigos também fossem amigos era basicamente a mesma de que eles conhecessem pessoalmente o Dalai Lama ou o presidente dos Estados Unidos. Quer dizer, em grafos aleatórios o coeficiente de agregação é extremamente baixo. Com o objetivo de contornar esta falha, o sociólogo Duncan J. Watts e o matemático Steven H. Strogatz, propuseram, em 1998, um modelo que interpola uma rede regular e uma rede aleatória [18], e assim conseguiram replicar, em um único modelo de redes, características de mundo pequeno e o alto coeficiente de agregação presentes nas redes reais.

Esse novo modelo falhava em reproduzir uma outra característica essencial de várias redes, a diversidade na conectividade, a qual mede a quantidade de conexões que os nós possuem. Tanto os modelos de Erdős-Rényi e de Watts-Strogatz prediziam que a conectividade seria descrita por uma distribuição de Poisson, implicando que a grande maioria dos nós teria uma conectividade próxima de um valor médio, com desvios que seriam exponencialmente suprimidos. No entanto, várias redes não são igualitárias, já que contém a presença de pólos (*hubs*), nós altamente conectados que caracterizam as chamadas redes livres de escala, já que a conectividade é descrita por uma lei de potência e não uma por uma conectividade média característica. Com o intuito de reproduzir esse comportamento livre de escala, introduziu-se o chamado modelo de Barabási-Albert [19], redescobrimto independente de uma outra lei de potência, proposta por Derek Price em 1965 para descrever as redes de citações entre artigos científicos [20]. O modelo de Barabási-Albert, se baseia em dois conceitos centrais: redes dinâmicas, implicando que a rede está em constante crescimento, e a ligação preferencial (chamada por Price de vantagem cumulativa em um artigo de 1976 [21]), dizendo que um novo nó conectado à rede existente tem uma probabilidade muito maior de se conectar a nós mais conectados do que àqueles com

poucas conexões, propriedade popularmente chamada de “o rico fica mais rico”, ou efeito Matthew [22].

Feita esta introdução histórica, o primeiro objetivo deste artigo será apresentar em mais detalhes cada um dos modelos de redes discutidos acima – Erdős-Rényi, Watts-Strogatz e Barabási-Albert – assim como as suas principais propriedades estatísticas tais como menor caminho médio, coeficiente de agregação e distribuição de conectividade. Nosso segundo objetivo é mostrar uma aplicação prática desses conceitos em uma rede real, para o qual escolhemos analisar a rede de pesquisadores do Instituto Nacional de Ciência e Tecnologia de Informação Quântica (INCT-IQ) [23]. Com este exemplo, não somente poderemos ilustrar os principais conceitos da ciência das redes, assim como demonstrar e discutir as características e o sucesso de uma importante rede de colaborações científicas do Brasil.

Nosso artigo é organizado da seguinte forma. Na Seção 2, revisamos as propriedades estruturais das redes, definindo os conceitos básicos e os modelos de redes. Na Seção 3, fazemos um breve histórico do INCT-IQ, para depois discutirmos os aspectos quantitativos na Seção 4. Na Seção 5, apresentamos as propriedades estatísticas desta rede, enquanto que na Seção 6, oferecemos diferentes formas de se visualizar a rede dos principais pesquisadores do INCT-IQ. Por fim, na Seção 8 apresentamos nossas conclusões.

2. Propriedades estruturais das redes

Redes são formadas por um conjunto V de vértices e um conjunto E de arestas, onde cada aresta é definida por um par de vértices (i, j) que representa algum tipo de relacionamento ou conexão entre tais vértices. Vértices que compartilham uma aresta são denominados vizinhos [24]. No caso de redes direcionadas, como a ilustrada na Fig. 2(b), as arestas possuem direções de modo que $(i, j) \neq (j, i)$; alguns exemplos desta característica são os hiperlinks entre páginas web e

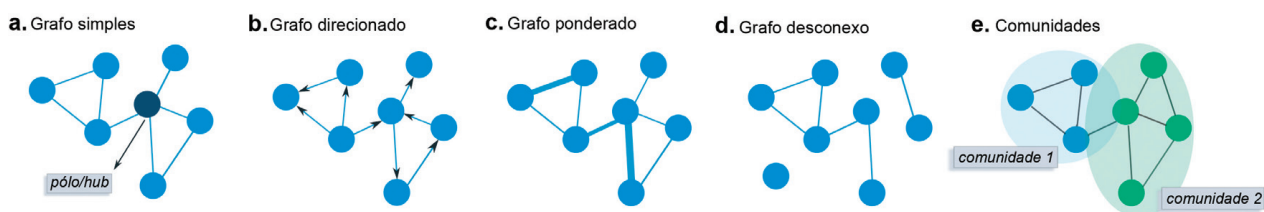


Figura 2: Exemplos de grafos, onde os círculos representam os vértices e as ligações entre eles correspondem às arestas. **a.** Grafo simples com 7 vértices e 8 arestas, em que cada vértice é ligado a outro com apenas uma aresta e um vértice não se conecta com ele mesmo (ausência de auto-loops). Vemos também a existência de um pólo, destacado com a cor azul escura, ou seja, o nó com o número maior de conexões quando comparado aos outros nós da rede. **b.** Exemplo de grafo direcionado (orientado), no qual as arestas possuem direções, isto implica que $(i, j) \neq (j, i)$ onde i e j são os vértices da rede. Nesta configuração os vértices possuem duas conectividades, k_{in} e k_{out} , que correspondem a conectividade de entrada e a conectividade de saída, respectivamente. **c.** Em grafos ponderados cada aresta (i, j) possui um peso que pode representar, por exemplo, o custo para ir um vértice a outro. Na figura, o peso de cada ligação está associado à largura da aresta. **d.** Representação gráfica de uma rede desconexa, no qual não existe um caminho que percorre todos os vértices da rede. **e.** Exemplo de uma rede com duas comunidades, observamos que existem dois grupos interconectados, onde a conexão entre eles é realizada pelo pólo da rede.

redes de citações científicas. Além disso, há diversas outras propriedades relacionadas aos vértices e arestas em um grafo, por exemplo, arestas ou vértices possuem pesos caracterizam as redes ponderadas, como mostrado na Fig. 2(c). Por meio desta representação, é possível extrair diversas propriedades estatísticas da rede, que serão detalhadas a seguir, tais como distribuição de conectividade, coeficiente de agregação e menor caminho médio.

2.1. Armazenamento de redes em memória

Existem várias formas de representar a estrutura de uma rede no computador. Cada uma delas demonstra ser útil para uma situação específica, possuindo vantagens e desvantagens. A escolha de cada representação está relacionada tanto ao tempo computacional quanto às posições de memória de cada uma delas. Algumas representações típicas são: matriz de adjacência, matriz de incidência, lista de adjacência, vetorial e estrela direta e reversa. Nesta seção abordaremos duas delas.

1. *Lista de adjacência*: essa representação corresponde a um conjunto de listas, em que cada lista é identificada por um vértice i que contém o conjunto de todos os vértices que compartilham uma aresta com i (todos os seus primeiros vizinhos). Portanto, essa estrutura é composta por uma lista de dimensão N que contém dois campos, a identificação do nó e um ponteiro para uma lista encadeada contendo seus vizinhos. Esta é uma das representações mais utilizadas, já que ela ocupa $\mathcal{O}(N + E)$ posições de memória, onde E é a quantidade de ligações na rede. A Figura 3(b) ilustra esta representação para uma rede não direcionada e não ponderada.

2. *Matriz de adjacência*: essa estrutura é representada por uma matriz $A = [a_{ij}]$ quadrada de ordem N que ocupa $\mathcal{O}(N^2)$ posições de memória. Cada vértice é identificado pelos índices das linhas e colunas, de modo que cada elemento da matriz é preenchido pela relação:

$$a_{ij} = \begin{cases} 1, & \text{se os vértices } i \text{ e } j \text{ estão conectados} \\ 0, & \text{caso contrário.} \end{cases}$$

A Figura 3(c) ilustra a representação da matriz de adjacência para uma rede não direcionada onde as ligações não possuem pesos.

2.2. Distribuição de conectividade

Usualmente a medida estatística mais estudada na ciência das redes é o grau do vértice, que corresponde ao número de primeiros vizinhos que um vértice possui e, consequentemente, representa o número de ligações que ele realiza. A soma do grau de todos os vértices da rede está relacionada ao número total de arestas por meio da

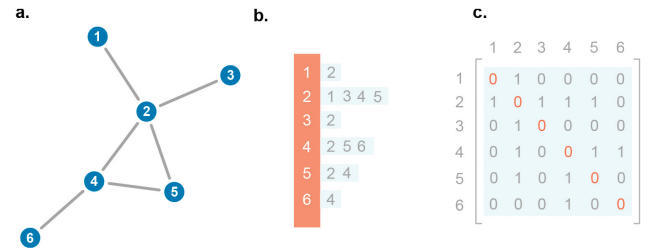


Figura 3: Armazenamento de redes em memória. **a.** Esquematização de uma rede com seis vértices e seis arestas com suas respectivas representações em **b.** Lista de adjacência e **c.** Matriz de adjacência. Observe que o grafo não possui múltiplas arestas entre o mesmo par de vértices e os elementos da diagonal da matriz de adjacência são nulos devido a não existir ligações que conectam um vértice a ele mesmo.

seguinte relação

$$\sum_{i=1}^N k_i = 2m, \quad (1)$$

onde N corresponde ao tamanho da rede, k_i é o grau do vértice i e m é o número total de arestas. Em algumas redes existem vértices que possuem um número muito grande de ligações quando comparados aos outros nós da rede. Neste caso, estes vértices são denominados de pólos (hubs).

Por meio da conectividade de cada vértice, é possível extrair duas propriedades estatísticas da rede. A primeira é o grau médio dos vértices, que consiste na média aritmética do grau de todos os vértices da rede, ou seja,

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i. \quad (2)$$

A segunda propriedade é a distribuição de conectividade, também denominada de distribuição de grau $P(k)$, é definida como a fração de vértices na rede que possui grau k , ou seja, é a probabilidade de um vértice, escolhido ao acaso, possuir k ligações. Caso a rede seja direcionada, os vértices possuem dois tipos de conectividade, a conectividade de entrada k_{in} e a conectividade de saída k_{out} que configuram as distribuições de conectividades $P(k_{in})$ e $P(k_{out})$, respectivamente.

Conforme discutido mais a frente, as redes de Erdős-Rényi e Watts-Strogatz tem sua distribuição de conectividade descrita por uma distribuição de Poisson, dada por

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \quad (3)$$

enquanto a rede de Barabási-Albert tem sua conectividade descrita por uma lei de potência, tal como

$$P(k) \propto k^{-\gamma}. \quad (4)$$

Ambas as distribuições são descritas na Figura 4. Redes com uma distribuição de grau da forma da

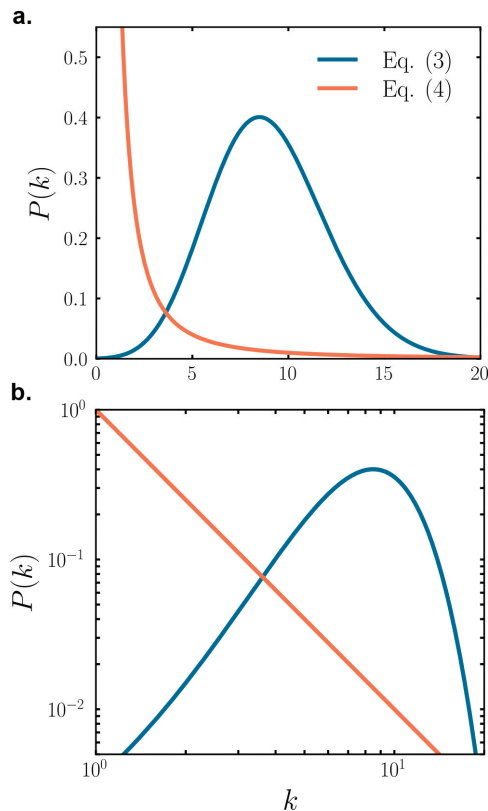


Figura 4: **a.** Comparação entre a distribuição de Poisson, Eq. (3), e a distribuição em Lei de Potência, Eq. (4), em escala linear-linear. Observamos que a distribuição em lei de potência possui uma cauda pesada, enquanto que a função exponencial decai mais rapidamente. **b.** Comparação entre as mesmas distribuições, porém em escala log-log. A distribuição lei de potência nesta escala corresponde a uma reta cuja inclinação é igual ao expoente γ .

Eq. (4) são denominadas de livres de escala, visto que as conectividades dos vértices não possuem um valor típico, em contraste com a Eq. (3), na qual a maioria dos vértices possuem conectividade próxima ao valor médio $\langle k \rangle$.

2.3. Distâncias e caminhos

A figura Fig. 5(a) ilustra o caminho entre dois vértices escolhidos aleatoriamente na rede, que é definido pela quantidade de arestas, sem repetições, que os conectam. Logo, se não há um caminho entre os sítios i e j , a distância entre eles será infinita, $d_{ij} = \infty$. Geralmente existe mais de um caminho que conecta um par de vértices na rede. No entanto, dois deles estabelecem duas importantes definições:

Menor distância: é o menor caminho, dentre todos os possíveis, que conecta dois vértices na rede. Por conseguinte, o menor caminho médio é dado pela média aritmética dos menores caminhos entre todos os $N(N-1)/2$ pares de vértices da rede. Matematicamente, para redes

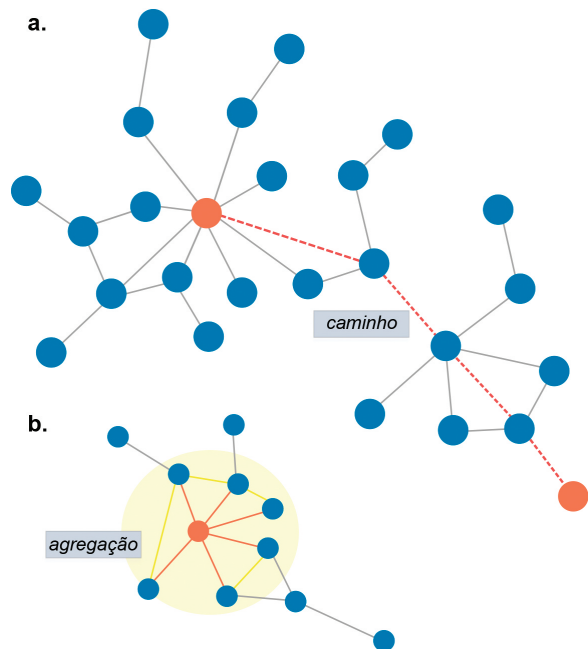


Figura 5: **a.** Representação ilustrativa de um caminho, dentre todos os possíveis, entre os vértices coloridos de laranja na rede. Note que, este caminho é o que contém a menor quantidade de ligações que separam os vértices, consequentemente, também é a menor distância entre eles. **b.** Representação do coeficiente de agregação do vértice laranja. Observe que, alguns dos vizinhos do vértice em destaque também estão conectados entre si (ligações em amarelo).

não ponderadas, esta quantidade é definida como

$$\langle l \rangle = \frac{2}{N(N-1)} \sum_{i < j} d_{ij}, \quad (5)$$

onde d_{ij} é a menor quantidade de arestas que conectam os vértices i e j . Note que nesta definição, somente os vértices que estejam conectados (excluindo-se os pares de distância infinita) são considerados. Para o caso de redes ponderadas, um caminho é definido como sendo a soma dos pesos de cada aresta entre os pares de vértices. Quando os pesos atribuídos às ligações são positivos é conveniente utilizar o algoritmo de Dijkstra [25], e para pesos que possam ser negativos o algoritmo de Ford-Moore-Bellman [26–28] é adequado para o cálculo do menor caminho médio.

Maior distância: a maior distância, dentre as menores distâncias, entre qualquer par de vértices caracteriza o chamado diâmetro da rede, definido com

$$d_{\max} = \max_{i,j \in V} d_{ij}. \quad (6)$$

Em alguns modelos de redes, tais como os modelos de Erdős-Rényi, Watts-Strogatz e Barabasi-Albert, a distância média dentre todos os caminhos da rede $\langle d \rangle$ escala como

$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}, \quad (7)$$

propriedade popularmente conhecido como mundo pequeno, na qual as distâncias características da rede (menor caminho médio ou o diâmetro) crescem logaritmicamente com o tamanho do sistema. Em redes de mundo pequeno, um par aleatório de vértices pode se conectar por meio de um pequeno número de arestas intermediárias, fato essencial para a eficiência, por exemplo a propagação de informação, de uma dada rede. O fenômeno do mundo pequeno é abordado em diversos estudos de redes complexas [29, 30], e até mesmo no meio cinematográfico [31]. Entretanto, já para redes regulares, onde cada vértice é conectado com a $\langle k \rangle$ vizinhos, o comprimento de menor caminho médio é proporcional a relação $\langle l \rangle \approx N/2\langle k \rangle$ [18]. Logo, esse tipo de rede não apresenta a propriedade de mundo pequeno.

2.4. Coeficiente de agregação

Outra importante propriedade para uma análise estatística em redes é o coeficiente de agregação (ou transitividade). Essa grandeza nos informa o quão agregados estão os primeiros vizinhos dos vértices na rede (ver Fig. 5(b)). Para um sítio i com conectividade k_i o coeficiente de agregação local é definido como [18]

$$c_i = \frac{2n_i}{k_i(k_i - 1)}, \quad (8)$$

em que n_i é o número de ligações entre os k_i vizinhos do sítio i e $k_j(k_j - 1)/2$ corresponde ao número total de ligações possíveis entre os vizinhos do sítio i . Esta propriedade é frequentemente estudada no campo das redes sociais, com a questão “Qual a probabilidade dos meus amigos também serem amigos entre si?” [32]. Note que, caso todos os vizinhos do sítio i também estejam conectados entre si, o valor do coeficiente de agregação local será $c_i = 1$, caso contrário, se nenhum dos vizinhos realizam conexões entre si obtemos $c_i = 0$.

O coeficiente de agregação médio da rede é obtido ao realizarmos a média aritmética sobre todos os coeficientes de agregação locais pelo tamanho da rede

$$\langle C \rangle = \frac{1}{N} \sum_i c_i. \quad (9)$$

Essa propriedade nos informa a probabilidade de que dois vizinhos de um sítio selecionado aleatoriamente estejam conectados entre si.

No caso das redes aleatórias geradas pelo modelo de Erdős-Rényi, o coeficiente de agregação é inversamente proporcional ao tamanho da rede $\langle C \rangle \sim 1/N$, valor pequeno e incompatível com a maioria das redes reais. Em contrapartida, a rede de Watts-Strogatz carrega a característica das redes regulares esparsas que possuem um alto valor para o coeficiente de agregação, de modo que este modelo captura tanto o efeito de mundo pequeno, como um alto valor para o coeficiente de agregação, $\langle C \rangle$ observado em redes reais. Finalmente, no modelo de Barabási-Albert, o coeficiente de agregação

é aproximadamente $\langle C \rangle \sim (\ln N)^2/N$, onde o termo $(\ln N)^2$ aumenta o valor desta quantidade. No entanto, à medida que $N \gg 1$, o coeficiente de agregação se aproxima de zero.

2.5. Aglomerados

Redes são classificadas conforme suas propriedades e dentre elas é importante estabelecer a definição de conexidade. Quando os vértices formam uma estrutura inteiramente conectada, de modo que sempre existe um caminho entre dois vértices aleatórios, identificamos que esta rede é conexa. Caso contrário, se existirem vértices isolados ou pequenos grupos de vértices conectados, classificamos a rede como desconexa (ver Fig. 2(d)). Intuitivamente, no caso de redes desconexas é possível calcular o número de aglomerados, onde aglomerados são definidos por subgrafos conexos e/ou vértices isolados dentro da rede. Logo, definimos o maior aglomerado N_G como sendo o subgrafo conexo com a maior quantidade de nós.

Nos modelos aleatórios de Erdős-Rényi e Watts-Strogatz, observamos a presença de alguns aglomerados na rede, como mostrado na Fig. 6, de tal forma que as redes não são completamente conectadas. Devido a isso, Newman e Watts propuseram uma modificação no modelo de Watts-Strogatz de modo que a rede contenha um único elemento conectado [33]. Redes geradas pelo modelo de Barabási-Albert possui apenas um único aglomerado conectado, sendo assim, elas são classificadas como conexas.

2.6. Assortatividade

O conceito do coeficiente de assortatividade foi introduzido por Newman [34], sendo calculada pela seguinte equação:

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}, \quad (10)$$

uma medida também conhecida como coeficiente de correlação de Pearson. Na equação acima vemos que estamos medindo a assortatividade com relação ao grau de conectividade dos nós, onde $-1 \leq r \leq 1$ e j_i, k_i são as conectividades dos vértices que compartilham uma ligação i , com $i = 1, \dots, m$, sendo m o número total de ligações. A assortatividade r nos informa a correlação entre sítios vizinhos (sítios que compartilham uma ligação). Dito de outra forma, esse coeficiente reflete a afinidade dos nós em se conectar com outros nós com os quais tenham alguma similaridade.

Caso tenhamos $r > 0$, a rede é dita ser assortativa e os vértices compartilham uma ligação com outros vértices que possuem em média o mesmo número de vizinhos. Caso contrário, se $r < 0$, a rede é dita disassortativa. Exemplos de redes assortativas incluem as redes de

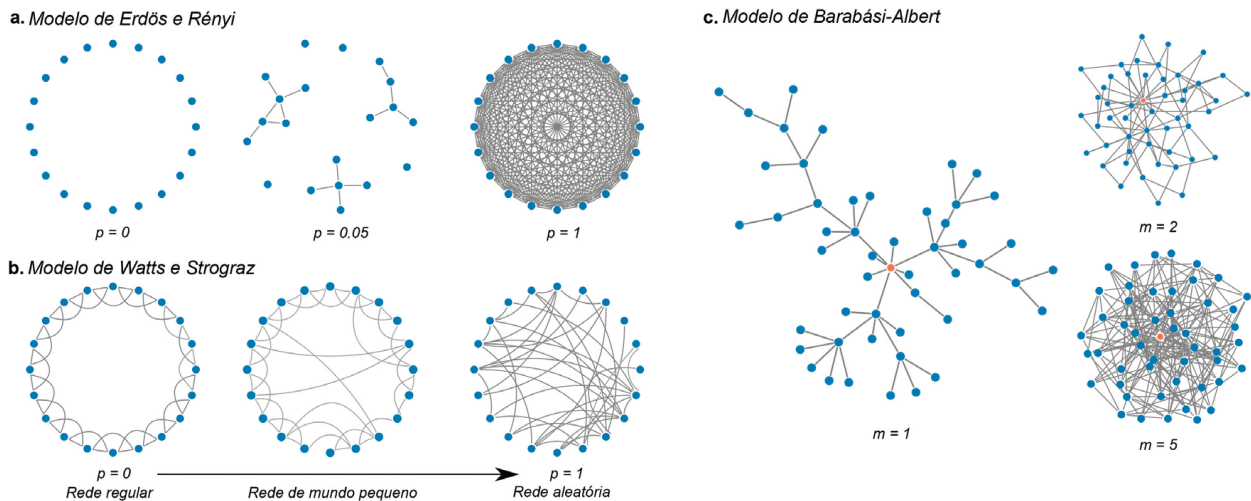


Figura 6: Exemplos de redes criadas por diferentes modelos matemáticos. **a.** Ilustração da rede para o modelo de Erdős e Rényi para típicos valores de p . Iniciamos a rede com $N = 20$ vértices isolados, logo em seguida, conectamos cada par de vértices (i, j) com probabilidade p . Observamos que para $p = 0$ a rede é composta por vários aglomerados de tamanho $N_G = 1$ e $\langle k \rangle = 0$, enquanto que para $p = 1$ obtemos uma rede conexa com $N_G = N$ e $\langle k \rangle = N - 1$. **b.** Representação do modelo de Watts e Strogatz com $N = 20$ vértices e inicialmente $\langle k \rangle = 4$ para três diferentes valores do parâmetro $p = [0, 0.09, 1]$. Notamos que quando $p = 0$ obtemos uma rede regular onde cada vértice possui k vizinhos. À medida que p aumenta, a rede torna-se cada vez mais aleatória, e quando $p = 1$, todas as ligações são reescritas de forma aleatória. No entanto, observamos que existe um regime em que a rede exibe as propriedades de redes de mundo pequeno. **c.** Exemplos de redes para o modelo de Barabási e Albert com $N = 50$ vértices e $m = [1, 2, 5]$, note que quando $m = 1$, a rede corresponde a um grafo tipo árvore. Contrariando os modelos de redes aleatórias de Erdős-Rényi e Watts-Strogatz, o modelo de Barabási-Albert exibe a propriedade de redes livres de escala, onde as conectividades dos vértices podem variar entre m até $N - 1$.

relações entre estudantes e redes de atores [34], enquanto a world-wide web e interações protéicas são exemplos de redes dissortativas [35].

2.7. Comunidades

Comunidades são definidas por nós bastante interconectados enquanto as conexões entre diferentes comunidades tendem a ser mais escassas (ver Fig. 2(e)) [36]. No caso de uma rede de colaborações científicas, uma comunidade representa um grupo de cientistas que trabalham e publicam juntos com certa frequência.

O método mais comum, chamado de método de Louvain, é baseado na otimização local da chamada modularidade de Newman-Girvan [37], que no caso de redes não ponderadas (como é o caso da rede que analisaremos aqui), é definida como

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \gamma \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \quad (11)$$

onde m é o número de conexões, A é a matriz de adjacência da rede, k_i é o grau de conectividade do nó i , γ é um parâmetro de resolução a ser escolhido, c_i é a comunidade a qual o i -ésimo nó da rede é associado e $\delta(c_i, c_j)$ é a função delta de Kronecker,

$$\delta(c_i, c_j) = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{caso contrário.} \end{cases}$$

Mais precisamente, a otimização da modularidade segue dois passos sequenciais básicos. O primeiro passo assume que cada nó da rede está associado a uma diferente comunidade, isto é, inicialmente o número de comunidades é igual ao número de nós da rede. Assim, dado um nó i , o algoritmo computa o aumento da modularidade, Eq. (11), se i é removido de sua comunidade e colocado em alguma das comunidades dos seus nós vizinhos. Se não há aumento na modularidade, o nó permanece em sua comunidade original, caso contrário é realocado na comunidade que maximiza (11). As comunidades geradas neste primeiro passo são chamadas de supervértices, tal que dois supervértices são conectados se existe ao menos uma conexão entre os nós individuais que compõem cada comunidade.

No segundo passo, a troca de comunidade de um dado nó continua, visando aumentar a modularidade, mas agora considerando não apenas nós vizinhos, mas também as conexões entre os supervértices. Claramente, o número de comunidades na rede pode apenas decrescer a cada interação do algoritmo, a menos que a modularidade atinja um valor máximo, situação na qual o processo de otimização é interrompido.

2.8. Modelos de redes

Redes são abundantes na natureza, consequentemente muitos pesquisadores propõem modelos teóricos que reproduzam, aproximadamente, as características de

alguns sistemas reais. Através de modelos teóricos podemos compreender a estrutura e prever alguns comportamentos de sistemas conectados. Nesta seção apresentaremos e discutiremos algumas das propriedades estruturais de três modelos de redes paradigmáticos: o modelo de rede aleatória de Erdős e Rényi (ER), o modelo de rede de mundo pequeno de Watts-Strogatz (WS) e o modelo de rede sem escala típica de Barabasi-Albert (BA).

2.8.1. Erdős-Rényi

No início da década de 1960 dois matemáticos, Paul Erdős e Alfréd Rényi, propuseram um modelo de rede aleatória [38], no qual um número fixo de vértices estão conectados aleatoriamente com a mesma probabilidade p . Dentro dessa estrutura, o modelo é construído em duas etapas:

1. A rede é iniciada com um conjunto de N vértices isolados.
2. Para cada par de vértices (i, j) sorteamos uma probabilidade aleatória de uma distribuição uniforme r , de modo que $0 \leq r \leq 1$. Caso $r \leq p$ os vértices i e j são conectados, caso contrário, a ligação não é criada.

Em uma rede aleatória, os graus dos vértices seguem uma distribuição de Poisson (Eq. 3). Isto indica que a maioria dos vértices possuem conectividade aproximadamente igual ao grau médio da rede, quer dizer

$$\langle k \rangle = \frac{2m}{N} = p(N-1) \simeq pN, \quad (12)$$

onde m é o número de ligações. Os casos extremos correspondem a $p = 0$, em que $m = 0$ (não há arestas) e $p = 1$ implicando que $m = N-1$, isto é, uma rede conexa onde cada vértice realiza $(N-1)$ conexões. Esta análise simples indica a existência de uma transição de fase entre uma rede desconectada para uma rede completamente conectada (ver Fig. 6(a)). Como resultado, observamos que a densidade de vértices que fazem parte do maior aglomerado varia de $N_G/N = 0$ para $N_G/N = 1$ a medida que aumentamos a probabilidade de conexão p . Entretanto, esta transição não ocorre gradualmente: existe um valor de probabilidade crítica p_c em que um componente gigante emerge. Erdős e Rényi mostraram que essa probabilidade está associada a condição $\langle k \rangle = 1$, de modo que,

$$p_c = \frac{1}{N-1} \approx \frac{1}{N}, \quad (13)$$

ou seja, um componente gigante irá existir se e somente se cada vértice possuir em média mais de uma ligação.

Como para o modelo ER, o coeficiente de agregação é equivalente à probabilidade de que dois vértices escolhidos aleatoriamente estejam conectados, obtemos que

$$\langle C \rangle = p = \frac{\langle k \rangle}{N}, \quad (14)$$

um comportamento inverso com o tamanho da rede e que é incompatível com os valores observados na maioria das redes reais. No entanto, uma propriedade importante no modelo de Erdős e Rényi é o efeito de mundo pequeno, de modo que o menor caminho médio é expressado pela Eq. (7). Este resultado indica que a menor distância média entre dois vértices em uma rede real de tamanho N é semelhante ao menor caminho médio de redes aleatórias de mesmo tamanho.

2.8.2. Watts-Strogatz

Duncan Watts e Steven Strogatz propuseram uma extensão [18] do modelo de redes aleatórias motivados por duas características estruturais comuns observadas em várias redes reais, sendo elas a propriedade de mundo pequeno e o alto coeficiente de agregação. Como vimos acima, no modelo de rede aleatória de Erdős e Rényi, o menor caminho médio entre dois vértices depende logarithmicamente com o tamanho da rede. No entanto, o coeficiente de agregação é muito baixo quando comparado aos valores observados em várias redes na natureza. Sabendo que redes regulares possuem um alto valor para o coeficiente de agregação, juntamente com a característica de mundo pequeno observada no modelo de rede aleatória de Erdős e Rényi, Watts e Strogatz formularam um modelo de rede aleatória que transita entre uma rede regular e uma rede aleatória, chamando-a de *small world network*, no português, rede de mundo pequeno. O modelo de Watts e Strogatz é definido pelas seguintes etapas:

1. Iniciamos a rede com N vértices conectados aos seus k primeiros vizinhos (círculo regular).
2. Com uma probabilidade p reposicionamos de forma aleatória e uniforme cada aresta da rede, de modo que auto conexões e conexões duplas não são permitidas.

Em uma rede regular o coeficiente de agregação é alto, assim como as distâncias na rede. Por outro lado, redes aleatórias apresentam um baixo coeficiente de agregação e a característica de mundo pequeno. Porém, é importante notar que a transição entre o regime de uma rede regular e uma rede completamente aleatória não ocorre bruscamente: uma transição suave ocorre entre os dois regimes. Conforme p aumenta, é possível observar o surgimento de atalhos que criam conexões de longo alcance, o que reduz as distâncias entre alguns pares de vértices. Ainda assim, esta mudança na configuração da rede não influencia grandemente a rede inicial, o que preserva o alto valor do coeficiente de agregação. Com isso, notamos que existe um intervalo de parâmetros p onde a rede apresenta o fenômeno de mundo pequeno e um alto coeficiente de agregação (ver Fig. 6(b)). Consequentemente, o comprimento de menor caminho médio é aproximadamente,

$$\langle l \rangle \sim \frac{1}{p} \ln(Np). \quad (15)$$

2.8.3. Barabási-Albert

Nos modelos de redes aleatórias apresentados anteriormente, nota-se que a distribuição da conectividade dos vértices é descrita por uma distribuição de Poisson. No entanto, observamos que muitas redes reais possuem uma distribuição de conectividade descrita, ao menos aproximadamente, por uma lei de potência, Eq. (4). Isto implica que existe uma grande variedade na conectividade dos nós, no qual eles não se limitam a possuir um valor próximo ao grau médio da rede, sendo esta propriedade denominada livre de escala. Este resultado indica que, diferente dos modelos de redes aleatórias, em redes livres de escala há o surgimento de pólos, vértices que contêm um número excepcional de ligações.

Albert-László Barabási e Réka Albert estudaram a rede da World Wide Web, e em seguida vários outros bancos de dados, e observaram que deveriam incluir duas características fundamentais, que são observadas em várias redes reais, na criação de um novo modelo de rede [12], são elas, (a) crescimento: o número de vértices cresce continuamente devido à adição de novos nós à rede, e (b) ligação preferencial: a interação entre dois vértices não é completamente aleatória, existe um mecanismo que determina quais vértices possuem mais chances de receberem conexões. O modelo proposto por Barabási e Albert [19] é assim construído em três etapas:

1. A rede é iniciada com um número pequeno de vértices m_0 , as ligações entre eles são escolhidas arbitrariamente, entretanto, cada nó precisa ter, no mínimo, uma ligação. Os vértices iniciais não podem se conectarem com eles mesmos (auto-loops) e não existem arestas múltiplas.

2. A cada passo de tempo t um novo vértice é adicionado à rede que se liga preferencialmente a $m(\leq m_0)$ outros vértices pré-existentes no sistema. A probabilidade de um vértice pré-existente na rede i receber a nova ligação é proporcional à sua conectividade,

$$\prod(k_i) = \frac{k_i}{\sum_j k_j}, \quad (16)$$

no qual sorteamos uma probabilidade aleatória de uma distribuição uniforme r , de modo que $0 \leq r \leq 1$, e o vértice pré-existente na rede i que receberá a ligação obedece a desigualdade $r \leq \prod(k_i)$.

3. Repete-se os passos anteriores até o tamanho desejado do sistema, após t passos. Este procedimento resulta em uma rede contendo $N = (m_0 + t)$ vértices e $(m_0 + mt)$ ligações, onde m é o número de conexões que cada vértice realiza ao ser inserido no sistema.

Neste modelo, a ligação preferencial está relacionada somente à conectividade do vértice, implicando que os nós mais conectados na rede tendem a receber mais conexões, gerando a formação de pólos. Desse modo, a idade do vértice é um fator determinante para ele se tornar ou não um pólo, visto que é mais provável que o nó inserido no início da criação da rede obtenha mais

ligações do que os adicionados posteriormente. Observe que, na Figura 6(c) muitos vértices na rede possuem poucas ligações, enquanto alguns deles tornam-se pólos e possuem várias conexões.

Barabási e Albert mostraram numericamente e analiticamente, utilizando a teoria contínua, que a distribuição de conectividade $P(k)$ deste modelo é descrita por uma lei de potência com um expoente fixo $\gamma = 3$, onde este expoente independe do parâmetro m e do tempo t .

O coeficiente de agregação do modelo Barabási-Albert para $m = 1$ é nulo, visto que nesta situação a rede corresponde a um grafo do tipo árvore, Fig. 6(c), onde o número de ligações equivale a $(N - 1)$. Em contrapartida, para $m > 1$ o coeficiente de agregação decai com o tamanho da rede, tal que

$$\langle C \rangle \sim \frac{(\ln N)^2}{N}, \quad (17)$$

onde o termo $(\ln N)^2$ aumenta esta quantidade. Isto significa que, os vértices no modelo de Barabási-Albert são localmente mais agregados do que os vértices nos modelos de redes aleatórias, quando comparamos para o mesmo tamanho de rede N .

De maneira equivalente, o comprimento de menor caminho para o modelo de Barabási-Albert é dividido nas seguintes situações, para $m = 1$ temos que [39]

$$\langle l \rangle \sim \ln N, \quad (18)$$

enquanto que, para $m > 1$,

$$\langle l \rangle \sim \frac{\ln N}{\ln \ln N}, \quad (19)$$

de modo que ambas as situações apresentam o efeito de mundo pequeno.

3. Breve histórico do INCT-IQ

A informação quântica é uma área de pesquisa multidisciplinar, iniciada ainda na década de 80 com os trabalhos precursores tal como a ideia de simulação quântica proposta por Feynman [40] e a versão quântica da tese de Church-Turing capitaneada por Deutsch [41]. A informação quântica evoluiu rapidamente ao longo da década seguinte, se tornando uma das áreas mais ativas e com maior crescimento nos dias de hoje, fato que se deve especialmente as variadas promessas de tecnologias quânticas: simulações quânticas eficientes [42], computação quântica [43], sensores quânticos [44], comunicação quântica [45] e criptografia quântica [46].

O Brasil contribui desde o surgimento desta área, com colaborações de pesquisadores no país iniciadas na década de 90. Estas colaborações foram fortalecidas a partir da implementação do programa “Institutos do Milênio” que surge em 2001 como uma iniciativa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [47]. O objetivo central era estabelecer

redes de pesquisa de padrão internacional em áreas de fronteira e consideradas de interesse estratégico, com um investimento de 90 milhões de reais provindos do extinto Ministério de Ciência e Tecnologia (MCT)[48]. No primeiro edital, foram recebidas 206 propostas das quais apenas 17 foram implementadas, uma das quais sendo o Instituto de Informação Quântica, coordenado pelo Prof. Dr. Luiz Davidovich do Instituto de Física da Universidade Federal de Rio de Janeiro (IF-UFRJ) e que mais tarde viria a se tornar presidente da Academia Brasileira de Ciências (ABC). Na época, a rede de colaboração estabelecida era formada por dez instituições em cinco estados do país, das regiões Sudeste e Nordeste, sendo a UFRJ a instituição sede. No estado do Rio de Janeiro, a rede também incluía o Centro Brasileiro de Pesquisas Físicas (CBPF) e a Universidade Federal Fluminense (UFF). No estado de São Paulo, faziam parte do Instituto pesquisadores do Laboratório Nacional de Luz Síncrotron (LNLS), além de professores da Universidade Estadual de Campinas (UNICAMP), da Universidade Federal de São Carlos (UFSCar) e da Universidade de São Paulo (USP), tanto do campus da cidade de São Paulo, quanto do Instituto de Física da cidade de São Carlos. A Universidade Federal de Minas Gerais (UFMG) era a única instituição mineira na rede, e na região Nordeste, o projeto contava com pesquisadores da Universidade Federal de Pernambuco (UFPE) e da Universidade Federal de Alagoas (UFAL) [49].

Em 2009, com o intuito de aumentar os recursos dedicados à esta exitosa iniciativa, surgem os Institutos Nacionais de Ciência e Tecnologia (INCTs), através de um edital com recursos compostos do CNPq e das Fundações de Apoio à Pesquisa (FAPs), agências de fomento estaduais existentes em alguns estados brasileiros e no Distrito Federal [50]. O aumento de recursos financeiros levou ao considerável crescimento do programa, que contratou 122 projetos na chamada de 2008, dentre estes o INCT-IQ, coordenado pelo Prof. Dr. Amir Caldeira (UNICAMP). A rede de 2008 contava com 66 pesquisadores principais agora de 8 estados brasileiros (Alagoas, Ceará, Espírito Santo, Minas Gerais, Paraná, Pernambuco, Rio de Janeiro e São Paulo) em três regiões do Brasil (Nordeste, Sudeste e Sul) [51].

O lançamento de uma nova chamada [52] para os INCTs em 2014 permitiu a continuação da iniciativa. O atual INCT-IQ [23], cujas atividades se iniciaram de fato em 2016, tem como instituição sede a Universidade Federal do Rio de Janeiro, sob a coordenação da Profa. Dra. Belita Koiller, pesquisadora 1A do CNPq e membro da Academia Brasileira de Ciências. A rede se expande com relação às antecessoras, e conta atualmente com 120 pesquisadores de 4 regiões do Brasil (Nordeste, Centro-Oeste, Sudeste e Sul), 12 estados brasileiros (Rio Grande do Norte, Pernambuco, Alagoas, Mato Grosso, Goiás, Espírito Santo, Rio de Janeiro, Minas Gerais, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul) e 20 instituições no Brasil [53].

4. Aspectos quantitativos da rede INCT-IQ

Utilizando como fonte de dados a plataforma Lattes, obtemos informações a respeito dos pesquisadores principais do INCT-IQ, tais como: filiação, produtividade e artigos publicados no período entre 2016 até o dia 29 de abril de 2022. Extraímos alguns aspectos quantitativos como a frequência dos tópicos principais nos títulos dos artigos, número de colaborações por artigo, número de publicações por pesquisadores, assim como o número de publicações e o número de pesquisadores em cada cidade do Brasil.

Na Fig. 7 visualizamos a nuvem de palavras contendo os 80 termos mais utilizados nos títulos dos artigos publicados pelos pesquisadores do INCT-IQ. No total, analisamos 1156 títulos que somam 8474 palavras. Eliminando os termos de conjunção, observamos que *Quantum*, *Entanglement*, *State* e *System* são as palavras mais comuns com respectivamente 300, 69, 63 e 62 repetições cada.

Por sua vez, na Fig. 8(a) traçamos o número total de artigos com um determinado número de coautores que pertencem ao INCT-IQ como pesquisadores principais. A maioria dos artigos é formado por um único autor, logo em seguida, o número de artigos decai rapidamente para artigos que contém dois ou mais autores. Vale notar que isto não significa que a maioria dos artigos sejam

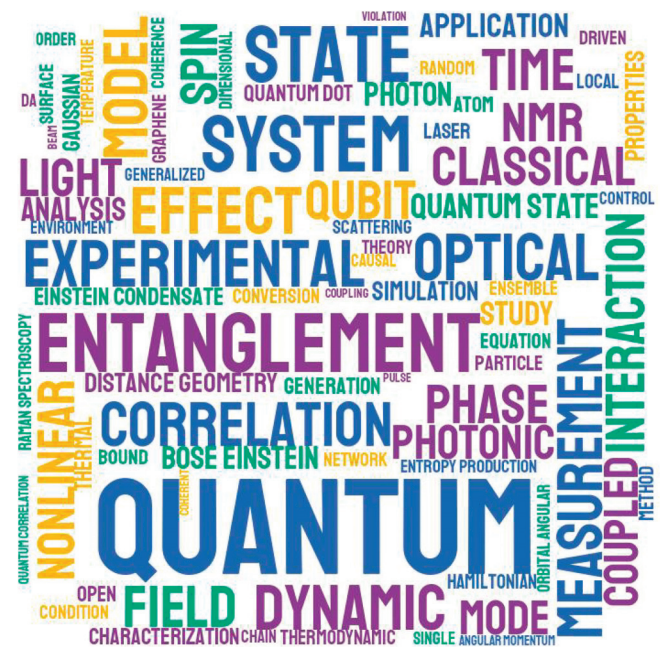


Figura 7: Visualização da nuvem de palavras formada pelos títulos dos artigos dos pesquisadores do INCT-IQ. Cada palavra possui um tamanho que depende da sua frequência no texto formado por todos os títulos. Observamos que, *Quantum*, *Entanglement*, *State* e *System* são os termos utilizados com maior frequência.

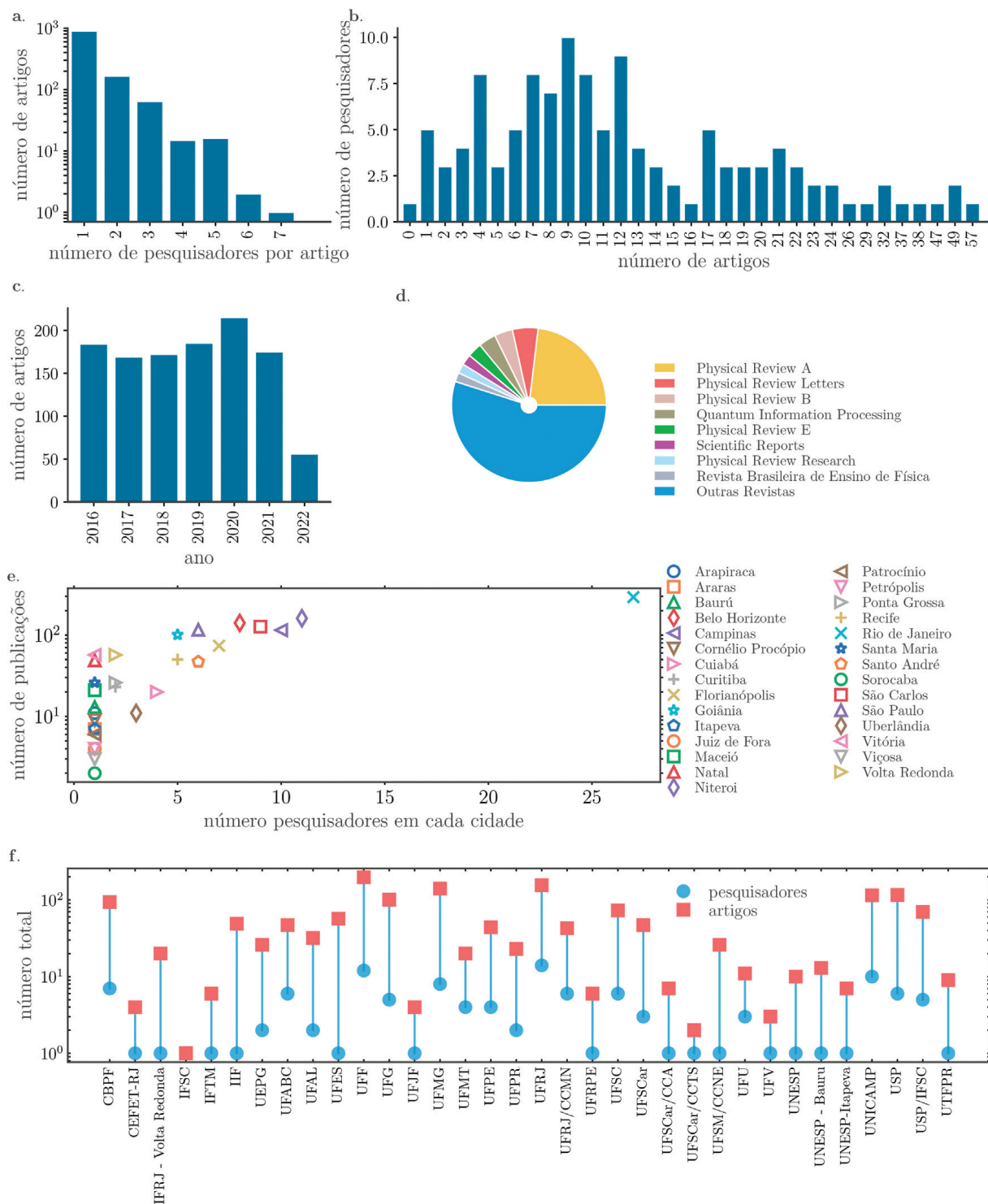


Figura 8: (a) Histograma do número total de artigos com um determinado número de coautores que pertencem ao INCT-IQ como pesquisadores principais. Visualizamos que a maioria dos artigos é composto por apenas um pesquisador principal do INCT-IQ, indicando artigos compostos por um único grupo de pesquisa. (b) Histograma do número de publicações por autor. Observamos que a maioria dos pesquisadores são autores de 9 ou 12 artigos. O pesquisador mais produtivo durante este período publicou 57 artigos. (c) Número total de artigos publicados pelos pesquisadores do INCT-IQ durante o período de 2016 a Abril de 2022. (d) Percentagem dos números de artigos publicados em revista específicas, onde mostramos as 8 revistas que possuem o maior número de artigos publicados pelos pesquisadores do INCT-IQ. Existem 267 revistas no total, em que observamos que a *Physical Review A* é a revista com a maior frequência, no entanto, existem 152 revistas com apenas uma única publicação. (e) Número de pesquisadores de cada cidade do Brasil em função do seu número cumulativo de publicações de cada região. Notamos que, o município do Rio de Janeiro é o que contém o maior número de pesquisadores, correspondendo também ao maior número de publicações. (f) Número de pesquisadores e de artigos publicados de cada Universidade/Centro no qual os pesquisadores do INCT-IQ atuam.

de autoria única, já que a maior parte dos co-autores dos artigos são colaboradores (por exemplo, alunos de pós-graduação) associados aos pesquisadores principais do INCT-IQ. Em nosso conjunto de dados identificamos apenas 1 publicação com 7 pesquisadores principais, que corresponde ao artigo da referência [54].

A Fig. 8(b) ilustra a relação do número de artigos que cada pesquisador possui no período de permanência no INCT-IQ. Observamos que o maior número de pesquisadores, 10 dentre um total de 120, possui autoria de 9 artigos, logo em seguida 9 cientistas são autores de 12 publicações. Por outro lado, apenas 5 colaboradores são autores de somente uma publicação. O nosso conjunto de dados mostra que o autor mais produtivo publicou 57 artigos desde sua adesão ao INCT-IQ.

Na Fig. 8(c) mostramos o número de artigos publicados em periódicos científicos pelos pesquisadores do INCT-IQ durante o período de Janeiro de 2016 a Abril de 2022. Observamos que, excluindo o número de publicações de 2022, o número de artigos publicado em cada ano durante o período de 2016 a 2021 é próximo ao valor médio de publicações, sendo este 183 artigos. É curioso notar que, apesar da pandemia do COVID-19, iniciada em 2020, este foi o ano com mais artigos publicados em periódicos científicos, com um total de 215 artigos publicados.

Considerando os artigos publicados pelos pesquisadores do INCT-IQ, a Fig. 8(d) mostra a proporção de artigos em oito revistas científicas. Observamos que, dentre a pluralidade de 267 periódicos, a *Physical Review A* é a que possui o maior número de publicações, logo em seguida, a *Physical Review Letters* ocupa a segunda posição. Além disso, 152 revistas específicas possuem apenas uma única publicação.

A Fig. 8(e) indica o número de pesquisadores de cada cidade do Brasil em função do seu número cumulativo de publicações de cada região. Verificamos que 29 municípios brasileiros possuem pesquisadores no campo da informação quântica do INCT, formando um total de 120 pesquisadores principais do INCT-IQ, do qual 30% residem no estado do Rio de Janeiro, onde a capital do estado homônimo e Niterói possuem 26 e 11 pesquisadores, respectivamente. Logo em seguida, Campinas abrange 10 cientistas. Em contrapartida, existem 14 cidades brasileiras que possuem somente 1 pesquisador do INCT-IQ. No entanto, dentre os municípios que correspondem ao único pesquisador, Vitória e Natal destacam-se pelo número considerável de publicações. Ainda, a Figura 8(e) mostra o número total de pesquisadores em cada Universidade/Centro brasileiro e o número total de publicações destes cientistas. Na Figura 8(f) realizamos a mesma análise vista em 8(e), porém contabilizando o número de pesquisadores e o número de publicações de cada Universidade/Centro no qual os pesquisadores do INCT-IQ atuam.

5. Propriedades estatísticas da rede INCT-IQ

A Tabela 1 apresenta algumas importantes propriedades estatísticas da rede de colaboração dos pesquisadores do INCT-IQ. Observamos que a rede possui $N = 120$ pesquisadores e um total de $E = 282$ colaborações, gerando 17 diferentes aglomerados na rede. Deste 17 aglomerados, 15 representam vértices isolados, quer dizer, pesquisadores que ainda não publicaram artigos em conjunto com outros participantes da rede. Outro aglomerado contém dois pesquisadores e o último e maior aglomerado concentra 85% dos pesquisadores, com $N_G = 103$, resultado que comprova que a rede de pesquisadores do INCT-IQ é de fato bastante interconectada.

Como discutido anteriormente, uma das propriedades estatísticas mais estudadas na ciência das redes é a distribuição de grau. Esta quantidade nos informa a fração de vértices na rede que possui determinado grau k , ou seja, no nosso cenário esta propriedade corresponde a probabilidade de um pesquisador escolhido ao acaso possuir k colaborações. Em muitas redes reais visualizamos que a distribuição de grau é ajustada por uma lei de potência, o que permite que alguns vértices possuam um número muito grande de ligações (denominados de pólos).

No caso da rede INCT-IQ, vemos que os dados são relativamente bem aproximados por uma distribuição de grau descrita por uma lei de potência truncada com um ajuste exponencial (ver Fig. 9),

$$P(k) = Ak^{-\gamma}e^{-\beta k} \quad (20)$$

onde A é a constante de normalização, $\gamma = 0.05$ é o parâmetro de decaimento e $\beta = 0.14$ controla a transição entre a lei de potência e o ajuste exponencial.

Tabela 1: Propriedades estatísticas da rede de colaboração dos pesquisadores do INCT-IQ.

Número total de artigos	1156
Número total de pesquisadores	$N = 120$
Número total de colaborações	$E = 282$
Média de autores principais do INCT-IQ por artigo	1.36
Média de publicações por pesquisadores	13.07
Número de aglomerados	17
Tamanho do maior aglomerado	$N_G = 103$
Segundo maior aglomerado	2 autores
Coefficiente de assortatividade	$r = 0.25$
Coefficiente de agregação	$\langle C \rangle = 0.38$
Menor caminho médio	$\langle l \rangle = 3.62$
Maior distância	$d = 9$
Média de colaboração por pesquisadores	$\langle k \rangle = 4.7$
Autor mais conectado	$k = 20$

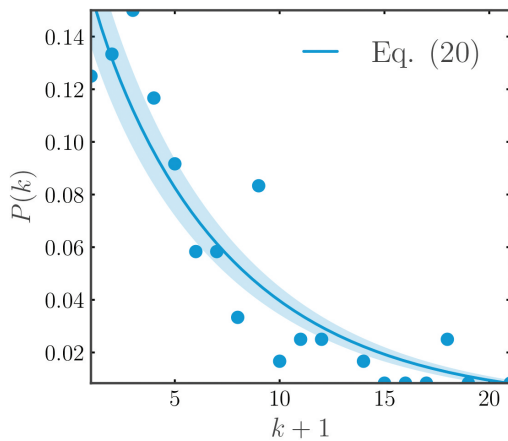


Figura 9: Distribuição de grau da rede de colaboração científica dos pesquisadores de informação quântica do INCT. Apesar das flutuações, é possível observar que a distribuição possui uma tendência similar a equação (20), que corresponde a uma lei de potência truncada por uma função exponencial. Este decaimento mostra que a maioria dos pesquisadores possuem poucas colaborações, enquanto que um pequeno grupo de pesquisadores realizam várias colaborações.

Esta análise nos mostra que a maioria dos pesquisadores do INCT-IQ realizam poucas colaborações entre si. No entanto, um pequeno número de pesquisadores possui várias colaborações. Embora tenhamos uma rede pequena, este resultado se alinha com outros estudos de redes de colaborações científicas [55] e interações sociais [56]. No último caso, em particular, vale citar o trabalho de Ribeiro e colaboradores [57], publicado na Revista Brasileira de Ensino de Física, analisando a rede de personagens da série de livros “Senhor do Anéis”, em que a distribuição de conectividade é bem ajustada pela equação (20) (considerando um $N = 618$).

Uma vez que identificamos o número de colaborações que cada pesquisador realiza, é possível calcular o número médio de colaborações da rede usando a expressão (2). Para o nosso conjunto de dados, encontramos que $\langle k \rangle = 4.7$, indicando que, em média, cada autor realiza entre 4 a 5 colaborações com outros pesquisadores do INCT-IQ. Ainda, observamos que o autor mais conectado realizou 20 colaborações com outros pesquisadores ao longo do período especificado.

Outra importante propriedade no estudo da ciência das redes são os seus caminhos, que correspondem ao número de arestas que separa dois vértices. Usando a expressão (5) encontramos que o menor caminho médio da rede é dado por $\langle l \rangle = 3.62$, no qual interpretamos que, em média, existem aproximadamente 4 pesquisadores separando quaisquer dois pesquisadores escolhidos aleatoriamente na rede. Além disso, observamos que a maior distância entre dois vértices do maior aglomerado é de apenas 9 ligações. Outra propriedade relevante para a análise estatística de redes é o coeficiente de agregação médio da rede $\langle C \rangle$, definido na Eq. (9). Esta propriedade mede a formação de grupos dentro da rede, de modo que

informa a probabilidade de dois cientistas colaborarem caso eles possuam um colaborador em comum. Os pesquisadores de informação quântica do INCT possuem um alto coeficiente de agregação $\langle C \rangle = 0.38$. Finalmente, calculamos o coeficiente de assortatividade, obtendo um valor de $r = 0.25$.

6. Visualização da rede

Com o objetivo de visualizar graficamente a rede de colaborações do INCT-IQ, utilizamos a biblioteca *NetworkX* [59] em Python para o cálculo das comunidades, onde esta biblioteca implementa o algoritmo de detecção de comunidades por Blondel et al. [60] utilizando a modularidade, dada pela Eq. (11). Quanto maior é o parâmetro γ maior serão as comunidades dentro da rede. Nós fixamos $\gamma = 1$, valor que nos retorna 8 comunidades. Para a representação da rede de colaboração, usamos o software *Gephi* com o layout circular, com ajustes manuais, particionado pelas comunidades. Na Fig. 11, visualizamos a rede de colaboração ao centro e ao lado todas as 8 comunidades separadas por cores. Observe que, a rede contém 17 aglomerados, sendo 15 deles formados por um único autor, ou seja, pesquisadores que ainda não realizaram colaborações. Dentro de cada comunidade identificamos a nuvem de palavras que correspondem aos principais temas de cada uma delas. Além disso, dentre os 120 pesquisadores da rede, somente 11 são mulheres.

Analisando as cidades e Universidades ou centros de pesquisa nas quais os pesquisadores se encontram, nota-se (ver Fig. 10) uma grande concentração no Sudeste brasileiro, que concentra 91 de um total de 120 pesquisadores. Curiosamente, 14 cidades contam com apenas um pesquisador. Analisando a rede de colaborações entre as cidades, notamos que existem 4 cidades que ainda não colaboraram entre si. Assim existe um maior aglomerado que contém 25 cidades com um grau de conectividade médio $\langle k \rangle = 5.8$, grau ponderado médio de conectividade dado por $\langle s \rangle = 19.45$ e onde os principais pólos são Rio de Janeiro e Niterói, com um número de conexões 168 e 74, respectivamente.

Para a rede de colaboração entre as cidades, a análise foi feita utilizando-se uma rede ponderada, em que cada ligação (i, j) corresponde ao número de colaborações entre os pesquisadores da cidade i com a cidade j . Além de atribuímos pesos nas ligações, caso um pesquisador da cidade i colabore com outro pesquisador que reside na mesma cidade, este vértice irá se conectar a ele mesmo, (auto-loop). Consequentemente, o grau ponderado, ou força, de cada vértice equivale a soma dos pesos das ligações que ele possui,

$$s_i = \sum_{j=1}^{k_i} w_{ij} \quad (21)$$

onde w_{ij} é o peso da ligação. Desse modo, o grau ponderado médio é a média sobre todos os s_i .

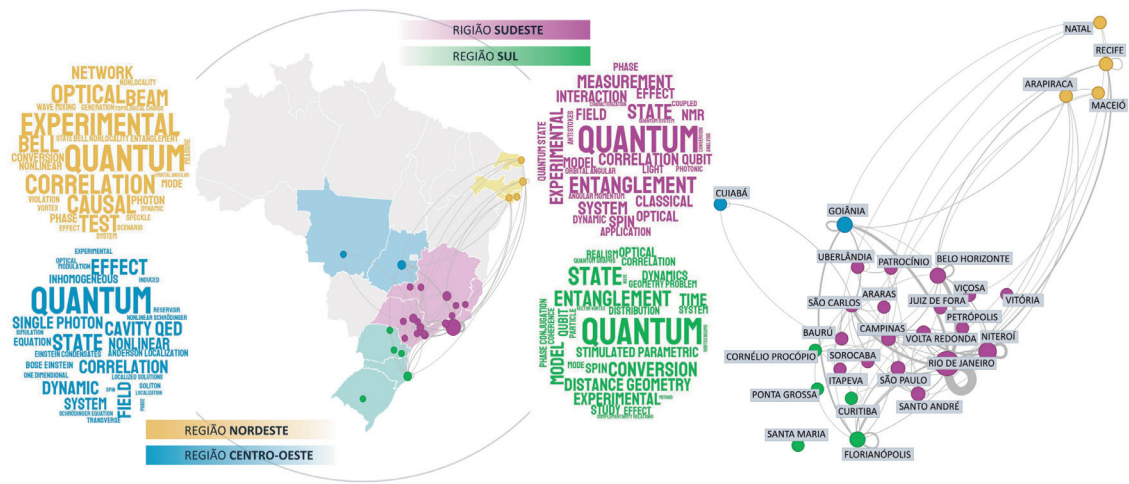


Figura 10: Rede ponderada de colaboração entre as cidades brasileiras. Caso um pesquisador da cidade A colabore com um pesquisador da cidade B, uma ligação é criada entre eles e o peso desta ligação corresponde ao número total de colaborações entre os pesquisadores que residem na cidade A com os pesquisadores que residem na cidade B. Portanto, cada colaboração acrescenta um valor unitário no peso da ligação, de modo que, na rede final, o peso das ligações corresponde ao número total de colaborações que cada cidade realiza entre si. Além de que, quando um autor colabora com outro da mesma cidade, o vértice da cidade se conecta a ele mesmo (auto-loop). Na ilustração, o tamanho do vértice é proporcional ao número de ligações que ele possui e a largura da aresta ao seu peso. Ainda na ilustração, observamos a nuvem de palavra de cada região brasileira que possuem pesquisadores do INCT-IQ.

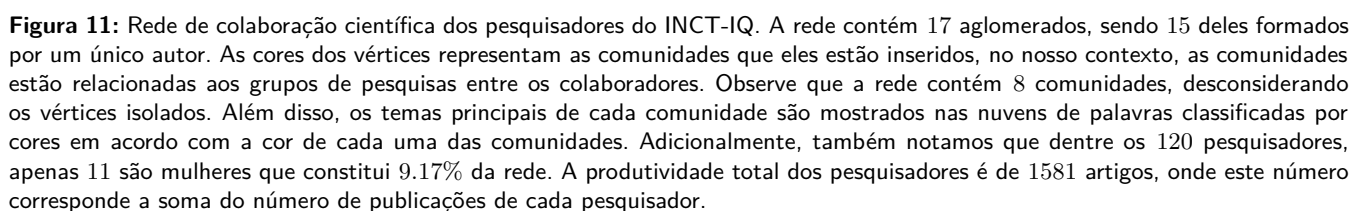
Tabela 2: Top 8 revistas que possuem o maior número de artigos publicados pelos pesquisadores do INCT-IQ. Além disso, à direita mostramos o número de publicações nas revistas de altíssimo fator impacto.

Revistas Com o Maior Número de Publicações	N.º de Artigos	Revistas de Altíssimo Fator Impacto	N.º de Artigos
Physical Review A	267	Nature	8
Physical Review Letters	62	Nature Communications	8
Physical Review B	44	Science	2
Quantum Information Processing	42	Nature Nanotechnology	2
Physical Review E	35	Science Advances	2
Scientific Reports	26	Nature Physics	1
Physical Review Research	23	Nature Photonics	1
Revista Brasileira de Ensino de Física	22	Physics Reports	1

7. Análise da rede do INCT-IQ

Nesta seção fazemos uma análise dos resultados encontrados dos nossos estudos da rede de pesquisadores do INCT-IQ. Com relação aos artigos produzidos, a boa qualidade da produção da rede é atestada pelo alto percentual (32.26%, correspondendo a um total de 373 artigos) de publicações nas revistas Physical Review A (QUALIS A2 CAPES na área de Astronomia/Física na classificação de periódicos 2013–2016), Physical Review Letters (QUALIS A1) e Physical Review B (QUALIS A2). Além disso, considerando revista de altíssimo impacto, tais como Nature, Science, Nature Communications e similares, temos 25 artigos no período analisado, como consta na Tabela 2. Os termos dos títulos dos trabalhos, ilustrados pela nuvem de palavras da Figura 7, mostram que a comunidade da rede se interessa majoritariamente pelo fenômeno do emaranhamento quântico, estudado tanto teórica quanto experimentalmente. O aglomerado principal, mostrado na Figura 11, se divide

em 8 comunidades, cujos os principais termos são: Anti-Stokes, Quantum Correlation, Quantum System, Quantum State, Quantum Effect, Quantum Orbital Angular e Quantum Experimental. Ao aprofundar o análise dos assuntos da produção desse aglomerado destacam-se artigos em revistas importantes na área nos seguintes assuntos: termodinâmica quântica, estados de momento angular, estudos teóricos de fundamentos de mecânica quântica e informação quântica em sistemas da física da matéria condensada. Nota-se um alto número de trabalhos na área de óptica, tanto de cunho experimental quanto teórico, o que contrasta com resultados prévios do análise das referências cruzadas de trabalhos em informação quântica disponíveis no repositório ArXiv com outras áreas, que mostra uma interseção majoritária com a área de física da matéria condensada [58]. Este fato se mostra uma característica própria da rede brasileira de pesquisadores em informação quântica. Analisando os números apresentados na Seção 5, o alto grau de interconectividade entre os participantes



assinala que os pesquisadores do INCT-IQ tendem a formar conexões com outros pesquisadores que possuem, aproximadamente, o mesmo número de colaborações entre si, fato similar ao que acontece na rede mundial de pesquisadores em informação quântica [58] ou mesmo na física como um todo [34].

A análise da interconectividade em termos geográficos mostra que considerando-se as 29 cidades envolvidas no INCT-IQ, nota-se que 25 destas formam o grande aglomerado de colaborações, sendo o Rio de Janeiro e Niterói seus principais pólos. Embora um conjunto de dados da rede do INCT-IQ correspondente ao período 2011–2014 não tenha sido analisada no presente trabalho, algumas informações extraídas da página oficial desse projeto

anterior [51] e do currículo Lattes dos pesquisadores principais nos permitem inferir alguns dados interessantes. A primeira é a expansão significativa da rede para fora dos grandes centros, passando de 12 em 2008 para 29 municípios brasileiros em 2016. Em particular, os pesquisadores das cidades de Cuiabá-MT, Florianópolis-SC, Goiânia-GO, Natal-RN e Santa Maria-RS, todos com produtividade destacada, desenvolveram suas teses de doutorado na região sudeste, nos estados de Rio de Janeiro e São Paulo, sob a orientação de membros dos projetos prévios (INCT-IQ/2008 e Institutos do Milênio). Estes pesquisadores são indicadores do fortalecimento da pesquisa da área no Brasil, pela sua capacidade de se engajar em colaborações nacionais diferentes daquelas do seu doutorado, o que fortalece o trabalho na área de informação quântica no país. A penetração da rede em cidades do interior do país, pela contratação de pesquisadores da área em campus avançados de Universidades ou Institutos Federais trazem uma capilaridade ausente das redes predecessoras. Defronte aos nossos dados, cabe afirmar que a existência da atual rede do INCT-IQ tem sido um motor por trás da manutenção da qualidade da produção dos pesquisadores lotados nessas instituições afastadas dos grandes centros de pesquisa.

8. Conclusões

Além de fazer uma introdução didática aos principais conceitos e modelos em ciência de redes, o objetivo principal deste artigo foi o de utilizar essas ferramentas para analisar as propriedades da rede de colaborações entre os cientistas do INCT-IQ.

Usando uma base de dados construída a partir dos nomes dos pesquisadores principais, pudemos inferir uma série de dados quantitativos sobre o INCT-IQ, sumarizados na Figura 8 e na Tabela 1. Do ponto de vista estatístico, apesar da rede ser consideravelmente pequena, a conectividade de seus nós é bem descrita por uma lei de potência truncada, em linha com o que foi observado em outras redes de colaborações [55]. Chama a atenção o alto grau de interconectividade entre os participantes do INCT-IQ, o maior aglomerado consiste de 85% dos cientistas, com um número médio de colaborações próximo de 5 e uma média de artigos publicados acima de 13 durante o período de duração do projeto. Destaca-se também o alto grau de agregação desta rede, $\langle C \rangle \approx 0.4$, mostrando que os colaboradores de um dado cientista também tendem a colaborar entre si. O aglomerado principal se divide em 8 comunidades, cujos os principais termos são: Anti-Stokes, Quantum Correlation, Quantum System, Quantum State, Quantum Effect, Quantum Orbital Angular e Quantum Experimental. Considerando-se as 29 cidades envolvidas no INCT-IQ, nota-se que 25 destas formam o grande aglomerado de colaborações, sendo o Rio de Janeiro e Niterói seus principais pólos.

Vale ressaltar que a partir da base de dados obtida, não pudemos incluir estudantes ou outros pesquisadores, que não aqueles que atuam como pesquisadores principais do INCT-IQ. Em análises futuras seria interessante que essa rede secundária fosse considerada, em particular, para se contabilizar o número de alunos formados ao longo dos anos. Também seria de valia contabilizar o impacto que as publicações do INCT-IQ vem conseguindo, por exemplo, através do número de citações alcançadas. Esperamos que este trabalho possa motivar estas e outras investigações futuras.

Agradecimentos

Agradecemos a Raíssa Zanela pela ilustração da Fig. 1 e a Samurá Brito pelo *script* em Python para a criação da rede de colaboração. Os autores agradecem o suporte financeiro do Instituto Serrapilheira (processo No. Serra-1708-15763), da Fundação Simons (processo No. 884966, AF), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), via o Instituto Nacional de Ciência e Tecnologia de Informação Quântica, processos No. 465469/2014-0 e 307295/2020-6, e das agências brasileiras MCTIC e MEC.

Referências

- [1] C. Kadushin, *Introduction to Social Network Theory* (MA, Boston, 2004).
- [2] O. Mason e M. Verwoerd, *Syst. Biol. IET* **1**, 89 (2007).
- [3] L.A. Meyers, B. Pourbohloul, M.E.J. Newman, D.M. Skowronski e R.C. Brunham, *J. Theor. Biol.* **232**, 71 (2005).
- [4] S. Brito, A. Canabarro, R. Chaves e D. Cavalcanti, *Phys. Rev. Lett.* **124**, 210501 (2020).
- [5] S. Brito, A. Canabarro, D. Cavalcanti e R. Chaves, *PRX Quantum* **2**, 010304 (2021).
- [6] N. Dehmamy, S. Milanlouei e A.L. Barabási, *Nature* **563**, 676 (2018).
- [7] V. Boginski, S. Butenko e P.M. Pardalos, *Comput. Oper. Res.*, **33**, 3171 (2006).
- [8] D. Petrone e V. Latora, *Sci. Rep.* **8**, 5561 (2018).
- [9] F. Caccioli, P. Barucca e T. Kobayashi, *J. Comput. Social Sci.* **1**, 81 (2018).
- [10] A. Sih, S.F. Hanser e K. A. McHugh, *Behav. Ecol. Sociobiol* **63**, 975 (2009).
- [11] T.G. Lewis, *Network science* (Wiley and Sons, New Jersey, 2009).
- [12] A.L. Barabási, *Network Science* (Cambridge University Press, Cambridge, 2016).
- [13] W.D. Wallis, *A beginner's guide to graph theory* (Springer Science & Business Media, Boston, 2007).
- [14] L. Euler, *Commun. Acad. Sci. Imper. Petropol.* **8**, 128 (1736).
- [15] P. Erdős e A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17 (1960).
- [16] F. Karinthy, in *The Structure and Dynamics of Networks*, editado por M. Newman, A.L. Barabási e D.J. Watts (Princeton University, New Jersey, 2006).

- [17] S. Milgram, *Psychol. Today* **1**, 60 (1967).
- [18] D.J. Watts e Steven H Strogatz, *Nature* **393**, 440 (1998).
- [19] A.L. Barabási e R. Albert, *Science* **286**, 509 (1999).
- [20] D. De Solla Price, *Science* **149**, 510 (1965).
- [21] D. De Solla Price, *J. Am. Soc. Inf. Sci.* **27**, 292 (1976).
- [22] R.K. Merton, *Science* **159**, 56 (1968).
- [23] <https://inctiq.if.ufrj.br>, acessado em 15/06/2022.
- [24] D.R. Figueiredo, in *Introdução a redes complexas*, editado por A.F. Souza, W. Meira Jr. (Editora PUC Rio, Rio de Janeiro, 2011).
- [25] W. Dijkstra, *Numer. Math.* **1**, 269 (1959).
- [26] L.R. Ford, *Network flow theory* (Rand Corp, Santa Monica, 1956).
- [27] E.F. Moore, *Proc. Int. Symp. Switching Theory*, **1959** 285 (1959).
- [28] R. Bellman, *Q. Appl. Math.* **16**, 87 (1958).
- [29] M.E.J. Newman, *J. Stat. Phys.* **101**, 819 (2000).
- [30] M. Kuperman e G. Abramson, *Phys. Rev. Lett.* **86**, 2909 (2001).
- [31] J. Gaure, *Six Degrees of Separation: A Play* (Vintage Books, New York, 1990).
- [32] S. Wasserman e K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994).
- [33] M.E.J. Newman e D.J. Watts, *Phys. Lett. A* **263**, 341 (1999).
- [34] M.E.J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [35] M.E.J. Newman, *Phys. Rev. E* **67**, 026126 (2003).
- [36] M. Girvan e M. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821 (2002).
- [37] M.E.J. Newman e M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
- [38] P. Erdős e A. Rényi, *Publ. Math. (Debrecen)* **6**, 290 (1959).
- [39] S. Dorogovtsev, *Complex networks* (Oxford University Press, Oxford, 2010).
- [40] R.P. Feynman, *Simulating physics with computers* (CRC Press, Flórida, 2018).
- [41] D. Deutsch, *Proc. R. Soc. London, Ser. A* **425**, 73 (1989).
- [42] S. Lloyd, *Science* **273**, 1073 (1996).
- [43] M.A. Nilesen e I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
- [44] C.L. Degen, F. Reinhard e P. Cappellaro, *Rev. Mod. Phys.* **89**, 035002 (2017).
- [45] N. Gisin e R. Thew, *Nat. Photon.* **1**, 165 (2007).
- [46] N. Gisin, G. Ribordy, W. Tittel, e H. Zbinden, *Rev. Mod. Phys.* **74**, 145 (2002).
- [47] Centro de memória do CNPq – Principais Realizações em 2001., disponível em: <https://centrodememoria.cnpq.br/realiz01.html>, acessado em 15/07/2022.
- [48] Programa Institutos do Milênio avança, disponível em: <https://revistapesquisa.fapesp.br/programa-institutos-do-milenio-avanca>, acessado em 15/07/2022.
- [49] Institutos do Milênio – Uma nova era para a pesquisa e o desenvolvimento do Brasil, disponível em: www.inesul.edu.br/site/documentos/instituto_milenio.pdf, acessado em 15/07/2022.
- [50] Página Oficial dos Institutos Nacionais de Ciência e Tecnologia, disponível em: <http://inct.cnpq.br/home>, acessado em 15/07/2022.
- [51] Instituto Nacional de Ciência e Tecnologia de Informação Quântica 2008-2013, disponível em: <https://sites.if.i.unicamp.br/inctiq>, acessado em 15/07/2022.
- [52] Chamada INCT – MCTI/CNPq/CAPES/FAPs nº 16/2014, http://inct.cnpq.br/documents/10180/124986/Chamada+INCT_16-2014.pdf/3d511440-8d6f-413c-ac64-176b7ac02902, accessed in 15/07/2022.
- [53] O INCT-IQ em números, disponível em: <https://inctiq.if.ufrj.br/inct-em-numeros>, acessado em 15/07/2022.
- [54] N.K. Bernardes, A.R.R. Carvalho, C.H. Monken e M.F. Santos, *Phys. Rev. A* **90**, 032111 (2014).
- [55] M.E.J. Newman, *Proc. Natl. Acad. Sci. USA* **98**, 404 (2001).
- [56] H. Ebel, L.I. Mielsch e S. Bornholdt, *Phys. Rev. E* **66**, 035103 (2002).
- [57] M.A. Ribeiro, R.A. Vosgerau, M.L. Andruchiw e S.E. Pinto, *Rev. Bras. de Ensino de Fis.* **38**, e3312 (2016).
- [58] S. Brito, R. Oliveira, R. Oliveira e R. Chaves, *arXiv*: 2112.03403 (2021).
- [59] A.A. Hagberg, D.A. Schult e P.J. Swart, in *Proceedings of the 7th Python in Science Conferenc*, editado por G. Varoquaux, T. Vaught e J. Millman (Pasadena, CA, 2008), p. 11.
- [60] V.D. Blondel, J.L. Guillaume, R. Lambiotte e E. Lefebvre, *J. Stat. Mech.* P10008 (2008).