

Título em Português: Identificando as Fronteiras da Física

Título em Inglês: Identifying Physics Frontiers

Autor: Matheus da Silva Fonseca

Instituição: Universidade de São Paulo

Unidade: Instituto de Física de São Carlos

Orientador: Luciano da Fontoura Costa

Área de Pesquisa / SubÁrea: Sistemas de Computação

Agência Financiadora: Sem fomento

Identificando as Fronteiras da Física

Matheus da Silva Fonseca

Luciano da F. Costa

Luciano da F. Costa

Universidade de São Paulo

math.sf.2019@usp.br

Objetivos

O estudo da ciência por redes complexas vem ocorrendo em larga escala e com diversos objetivos[1], tais como: entender padrões globais do desenvolvimento da ciência, estudar a integração de diferentes áreas de pesquisa e até formular grades curriculares de cursos de graduação.

Neste trabalho, inspirado em Silva et al.[2], buscamos modelar por meio de redes as subáreas das ciências físicas e utilizar a medida Acessibilidade para identificar suas fronteiras.

Com isso espera-se entender como se relacionam as diferentes subáreas, se elas possuem alguma intersecção (conhecimento em comum) e se essa ocorre através das fronteiras identificadas.

Buscamos também obter propriedades topológicas das redes que permitam entender a estrutura do desenvolvimento científico. Ademais, analisaremos se uma nova proposta de medida de rede que será chamada externalidade, definida como a razão entre número de nós na borda e nós totais da rede, se relaciona, de algum modo, com o sistema e sua dinâmica. Ainda tentaremos entender se a externalidade possui relação com medidas mais clássicas - Número de nós, Grau Médio, Coeficiente de Cluster e Menor Caminho Médio, por exemplo.

Ao final do trabalho espera-se uma compreensão melhor sobre a estrutura da ciência e como suas diversas áreas se desenvolvem e relacionam.

Métodos e Procedimentos

As redes serão criadas a partir de dados obtidos em bases cienciométricas como o arXiv ou ainda repositórios públicos como a Wikipédia em inglês. Os nós serão representados por artigos ou páginas e as arestas por citações, referências, entre outras. Utilizaremos a medida Acessibilidade que é calculada para cada nó das redes, dada uma certa dinâmica de movimento entre eles (tipicamente caminhada aleatória autoexcludente), e apresenta valores menores para nós mais fronteiriços. Calculada a acessibilidade podemos definir um *threshold* tal que nós com acessibilidade abaixo deles pertencem à borda e acima ao núcleo de rede. [3].

Identificada a fronteira, iremos calcular a nova medida proposta, chamada externalidade (b), dada pela equação 1, além de outras medidas clássicas de redes. Na equação 1, N é o número de nós da rede e N_b é o número de nós pertencentes à fronteira.

$$b = \frac{N_b}{N} \quad (1)$$

Observemos que esse valor representa um valor relativo do quanto da rede está na borda. Por ser uma medida nova, ainda estão em abertas possíveis interpretações para seus valores. Por exemplo, valores altos podem significar áreas em grande desenvolvimento. Este valor será comparado com as demais medidas através de métodos de estatística

multivariada como os Coeficientes de Pearson e Spearman e PCA. Iremos ainda comparar as medidas clássicas da borda com as do núcleo. Também será produzida uma rede reduzida em que cada nó representa uma das subáreas e suas ligações a intersecção entre elas. Essa será visualizada por um modelo novo, onde os nós se dividem em núcleo e borda e as arestas são de três tipos: núcleo - núcleo(nn), borda - borda(bb) e borda - núcleo(bn).

Resultados

Obtivemos até agora a rede de conhecimento com páginas da Wikipédia. Para isso foi utilizada uma divisão do próprio repositório em subcategorias de algum tópico (no caso, física) e obtivemos as subáreas.

Para essas redes foi obtida a acessibilidade para a dinâmica de caminhada aleatória autoexcludente de comprimento 2 e obtido os nós da borda com o *threshold* inicialmente igual à média da acessibilidade. A Figura 1 apresenta a rede da Mecânica Quântica com nós sem conexões removidos.

Desenvolvemos também a rede reduzida com as conexões nn e bb, Figura 2. Ainda foi feito um Scatter plot (Figura 3) em que os pontos representam arestas entre duas páginas, o eixo x é o peso da ligação bb, enquanto o y da nn e calculou-se o Coeficiente de Pearson dessas medidas, que foi de 0.55.

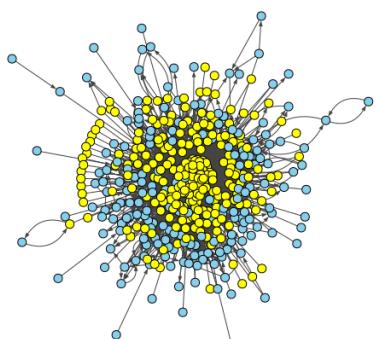


Figura 1: Rede da Mecânica Quântica.

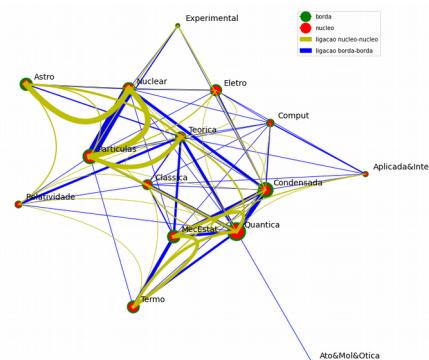


Figura 2: Rede Reduzida.

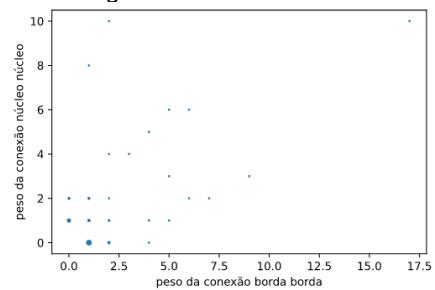


Figura 3: Scatter plot dos pesos das arestas. Eixo x: bb; Eixo y: nn.

Conclusões

A partir desses resultados preliminares concluiu-se que não existe uma relação muito forte entre borda-borda e núcleo-núcleo, no sentido do coeficiente de Pearson. Ademais, a quantidade de ligações bb e nn são próximas para o *threshold* adotado.

Referências Bibliográficas

- [1] COSTA, L. da F. et al. Characterization of complex networks: a survey of measurements. *Advances in Physics*, v. 56, n. 1, p. 167-242, Jan. 2007. DOI 10.1080/00018730601170527.
- [2] SILVA, F. N. et al. Identifying the borders of mathematical knowledge. *Journal of Physics A*, v. 43, n. 32, p. 325202-1-325202-7, Ago 2010. DOI 10.1088/1751-8113/43/32/325202.
- [3] TRAVENÇOLO, B. A. N.; VIANA, M. P.; COSTA, L. da F. Border detection in complex networks. *New Journal of Physics*, v. 11, n. 6, p. 063019-1-063019-17, June 2009. DOI 10.1088/1367-2630/11/6/063019.

Identifying Physics Frontiers

Matheus da Silva Fonseca

Luciano da F. Costa

Luciano da F. Costa

University of São Paulo

math.sf.2019@usp.br

Objectives

Complex networks studies of science are progressing at large scale and with diverse goals[1], like: understanding the global similarities in science development, studying the integration between different research areas and curricular design for undergraduate courses.

In this work, motivated by Silva et al.[2], we aim at modeling, by using complex networks, the subareas of physical sciences and use the Accessibility Measurement for identifying their frontiers.

In this way, we expect to better understand how these different subareas connect one another, if they have some intersection (common knowledge), and if this occurs along the identified frontiers.

We also aim at looking for topological properties of the networks which may allow us to better understand the structure of scientific development.

In addition, we will investigate if a new proposed measurement, that will be called externality, defined as the ratio between the numbers of borders nodes and total nodes of networks, is related in some way with the system and its dynamic. We will also try to understand if the externality has some relation with classical measurements – Number of nodes, Mean Degree, Cluster Coefficient and Average Shortest Path, for example.

Ultimately, we expect to achieve a better comprehension about the structures of science and how their areas develop and interrelate.

Materials and Methods

The networks will be created with data from scientometric bases, like arXiv or other public repositories like English Wikipedia. The nodes will be represented by articles or web pages while the edges reflect citation, references, or other possibilities.

We are using the Accessibility Measurement which is calculated for each node, given some dynamic on the network (normally self-avoided random walk), and has smaller values at border nodes. After calculating the accessibility, we can define a threshold such that a node with accessibility below it belongs to the border, while those exceeding it are part of the network nucleous.[3]

Next, we are going to calculate the proposed measurement, called externality, given by the equation 1, in addition to other classical network measurements. In equation 1 N is the total number of nodes in the networks and N_b is the number of nodes that belong to its frontier.

$$b = \frac{N_b}{N} \quad (1)$$

Observe that this value represents how many of the network nodes belong to the border. Being a new measurement, it is still open for possible interpretations. For example, higher values can mean more potential for further discoveries.

This measurement will be compared with other ones through methods of Multivariate Statistical analysis like Pearson and Spearman Coefficient and PCA. We will also compare the classical measurements of the border with those at the

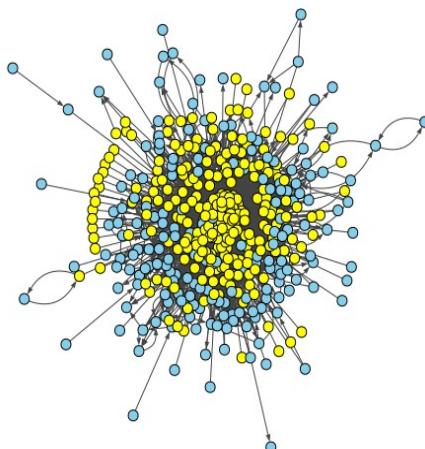
nucleo. Also, a reduced network can be derived in which the nodes and edges will represent each of the subareas and their intersection, respectively. This is going to be visualized in a possibly new manner, in which the nodes are divided in nucleous and border while the edges have three types: nucleo - nucleous (nn), border - border (bb) and border - nucleous (bn).

Results

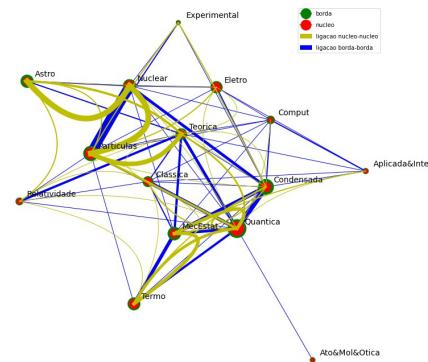
We obtained a preliminary version of the knowledge network for the Wikipedia pages. For that, we adopted the original subcategories of each topic (physics, in this case), and subareas.

For these networks, it was obtained the accessibility for the self-avoiding random walk dynamic of length 2 and the border nodes were determined with an initial threshold equal to the mean accessibility. Picture 1 presents the obtained Quantum Mechanics Network with zero-degree nodes removed.

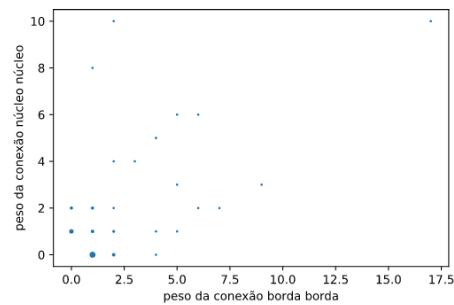
We also obtained the reduced network with the bb and nn connections, as shown in Picture 2. We also present a scatterplot (Picture 3) in which the points represent the edges between two pages, the x axis is the weight of the bb connection, while the y axis refers to nn. The Pearson Coefficient of these measurements was calculated as 0.55.



Picture 1: Quantum Mechanic Network.



Picture 2: Reduced Network.



Picture 3: Scatter plot of edges weights. x axe: bb; y axe: nn.

Conclusions

Considering these preliminary results it can be concluded that there is no big relationship between border-border and nucleo-nucleo, in the sense of Pearson Coefficient correlation. In addition, the amount of bb and nn edges are similar for the adopted threshold.

References

- [1] COSTA, L. da F. et al. Characterization of complex networks: a survey of measurements. *Advances in Physics*, v. 56, n. 1, p. 167-242, Jan. 2007. DOI 10.1080/00018730601170527.
- [2] SILVA, F. N. et al. Identifying the borders of mathematical knowledge. *Journal of Physics A*, v. 43, n. 32, p. 325202-1-325202-7, Ago 2010. DOI 10.1088/1751-8113/43/32/325202.
- [3] TRAVENÇOLO, B. A. N.; VIANA, M. P.; COSTA, L. da F. Border detection in complex networks. *New Journal of Physics*, v. 11, n. 6, p. 063019-1-063019-17, June 2009. DOI 10.1088/1367-2630/11/6/063019.