Introdução à renormalização em redes

Palavras-Chave: redes, grafos, renormalização, hiperbólica, simplificação

Autores:

Alan Albert Piovesana [Universidade Estadual de Campinas]

Prof. Dr. José Antônio Brum [Universidade Estadual de Campinas]

INTRODUÇÃO:

Este projeto está inserido no contexto do estudo de teoria de redes complexas. Redes complexas são grafos com grandes números de elementos (nós e arestas) e que apresentam estruturas topológicas particulares. A área tem diversas aplicações no mundo real, visto que podem-se representar todo tipo de sistema interagente como uma rede. Aplicações de redes vão desde sistemas biológicos, como o metabolismo celular [1] e a estrutura neural do cérebro, até redes colaborações científicas [2], dentre outras. Esse campo de estudo nasceu da teoria dos grafos, e adquiriu seu próprio caminho após os trabalhos de Watts e Strogatz [3] e de Albert e Barabási [4].

Dentro dessa área, o projeto focou em estudar a simplificação de redes com grandes números de nós e arestas. Esse é um tema de relevância em ciência de dados e pode ser abordado por diferentes métodos. O método escolhido aqui foi o de renormalização em redes. Renormalização é um conceito popularizado na física estatística por [5] e [6], e se refere a um processo que analisa como a troca de escala nos dados de um modelo afetam o modelo em si. No caso das redes, a troca de escala pode ser feita por alguma forma de agrupamento de nós que sejam considerados semelhantes.

Nesse contexto, foram estudados diversos métodos de renormalização em redes. Em particular, deu-se maior enfoque ao chamado método de renormalização geométrica, que é baseado na aplicação de fórmulas fechadas para renormalizar uma rede que esteja imersa em um espaço hiperbólico. Para isso, estudou-se como imergir redes reais no espaço hiperbólico para poder renormalizá-las em seguida, procurando-se validar intuitivamente o procedimento com a aplicação de outros algoritmos de simplificação de redes.

METODOLOGIA:

Para conduzir esse estudo, os seguintes passos foram realizados:

- Primeiro, estudaram-se os modelos de geração artificial de redes no espaço hiperbólico, denominados H2 e S1. Esses modelos imergem os nós da rede em um espaço métrico, de forma que eles adquirem posição fixa no espaço, baseado em 2 parâmetros considerados úteis para se descreverem redes complexas: a popularidade de cada nó e a similaridade entre pares de nós.
- Como apenas gerar redes artificiais que respeitem certos modelos não é suficiente para aplicações em redes do mundo real, estudou-se um método que mapeia redes reais para tais espaços, denominado método de Mercator [7]. Esse método foi aplicado para algumas redes conhecidas, obtidas no repositório online 8.
- Com as redes imersas no disco hiperbólico, pôde-se, então, proceder com o método de renormalização geométrica das mesmas. Esse método foi baseado nos artigos [9] e [10], e construiu-se uma rotina simplificada em Python 3 [11] para renormalizar redes já previamente imersas no espaço hiperbólico. Esse programa recebe como entrada a lista de arestas da rede e o número de passos de renormalização desejado e devolve, em uma rede multicamada, as diversas redes intermediárias geradas pela iteração do procedimento.
- Por fim, no artigo [10], os autores validam a aplicação do método de renormalização geométrica comparando as redes intermediárias geradas com camadas coletadas empiricamente, e observam uma correspondência satisfatória entre os tipos de procedimento. Como não temos acesso a dados empíricos multi-escala, uma forma de validar o procedimento é aplicando outros métodos de simplificação de redes que tenham um apelo mais intuitivo. O algoritmo escolhido para isso - o chamado Método de Louvain [12] - é da classe dos chamados métodos de detecção de comunidades.

RESULTADOS E DISCUSSÃO:

Primeiro, para poder escolher qual seria o melhor algoritmo de detecção de comunidades para ser utilizado para validar a aplicação do método de renormalização geométrica, realizaram-se comparações em termos de qualidade das partições obtidas e de tempo de rodagem para redes grandes. Os algoritmos utilizados na comparação foram: o método espectral de Newman [13], o método de Louvain [12] e o método de Leiden [14]. Rodando-se os três algoritmos em um computador com processador i5 de 7ª geração, 8Gb de memória RAM e arquitetura x64, obtiveram-se os seguintes resultados. É de se observar que as comparações feitas são dependentes de diversos parâmetros, e não foram feitas com todo o rigor necessário para um experimento de *benchmark* formal.

Network	Number of nodes	Number of links	Running time [ms]			
			Louvain (NetworkX)	Louvain (Igraph)	Leiden (Igraph)	Newman (Igraph)
Karate	34	77	2.41 ± 0.04	0.23 ± 0.03	0.5 ± 0.1	13.5 ± 0.6
C. Elegans	453	2025	$(6 \pm 1)10^1$	3.4 ± 0.3	3.4 ± 0.2	71 ± 1
Drosophila Melanogaster	1781	8911	$(34 \pm 5)10^1$	16 ± 2	17 ± 4	$(42 \pm 2)10^1$
IMDB	896305	3782447	$(7 \pm 1)10^5$	$(46 \pm 2)10^3$	$(29 \pm 2)10^3$	-

Tabela 1 - Comparação dos diferentes algoritmos de detecção de comunidades em termos de tempo de rodagem

Como pode ser visto nos testes rodados, os algoritmos que performaram melhor foram o de Louvain e o de Leiden, ambos utilizando a biblioteca Igraph [15] do Python 3. Em particular, o método de Leiden se destacou. Os dois algoritmos se saíram melhor também em termos da qualidade das partições obtidas, quando medida pela métrica de modularidade [2]. Apesar desse desempenho, o algoritmo escolhido foi o de Louvain da biblioteca NetworkX [16]. O motivo para tal escolha foi a simplicidade de se acessarem as redes intermediárias que o algoritmo gera, o que permite uma comparação mais direta com o método de renormalização geométrica.

Utilizou-se, então, a rede do clube de karatê de Zachary [17] imersa no espaço S1 com auxílio do algoritmo de Mercator, e aplicou-se, por um lado, o método de renormalização geométrica, e, por outro, o método de Louvain da biblioteca NetworkX, guardando-se todas as camadas intermediárias geradas por ele. As redes multicamada obtidas por meio desses dois métodos foram as seguintes:

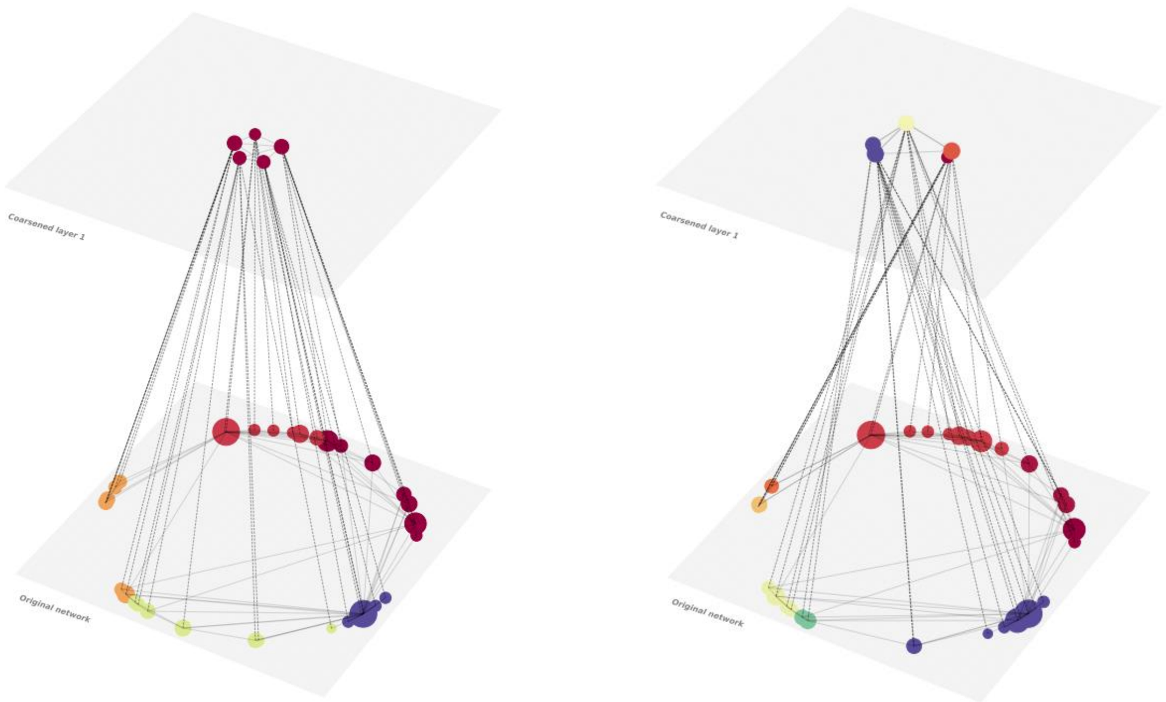

Figura 1 - Comparação entre a aplicação multicamada do método de renormalização geométrica (à esquerda) e do método de Louvain (à direita).

Como pode-se observar, para a rede do karatê, que possui 34 nós e 77 arestas, não há uma correspondência clara entre os dois métodos. No entanto, para redes maiores, e, em particular, redes obtidas experimentalmente, como no artigo [10], observou-se uma boa correspondência entre a renormalização geométrica e as redes empíricas.

Como próximos passos, deve-se testar o algoritmo em redes maiores e comparado com outros métodos de detecção de comunidades.

CONCLUSÕES:

Este projeto de iniciação científica com bolsa PIBIC e com vigência entre setembro de 2020 e setembro de 2021 foi focado em métodos de simplificação de redes complexas baseados em renormalização. Apesar de vários métodos alternativos de renormalização em redes terem sido estudados, o que ganhou maior enfoque foi o método de renormalização geométrica, baseado nos trabalhos 6, 7. Para que esse método seja aplicado, foi necessário que as redes estivessem imersas no disco hiperbólico, no modelo H2 ou no S1. Esses dois são modelos de geração de redes hiperbólicas artificiais, mas pode-se traduzir redes reais para o espaço hiperbólico utilizando o algoritmo de Mercator. Uma vez nesses espaços, realizou-se, iterativamente, a renormalização, gerando várias camadas intermediárias de rede até que se chegasse à escala desejada. Para essa aplicação da renormalização com caráter multicamada, construiu-se uma rotina em Python 3, presente no GitHub [11], que gera uma imagem completa de todas as redes hiperbólicas intermediárias geradas no processo. Por fim, comparou-se essa aplicação iterativa do algoritmo de renormalização com um método de detecção de comunidades denominado método de Louvain. Comparando-se os resultados camada-a-camada, não se observou correspondência entre as comunidades agrupadas para gerar redes em menor escala. Um dos motivos para essa não correspondência pode ter sido o fato de a rede testada (rede do clube de karatê de Zachary) apresentar poucos nós e arestas, estando muito longe do limite de redes grandes, que normalmente é assumido em teoremas de ciência de redes.

Como próximos passos deste projeto, pretende-se rodar o algoritmo em redes maiores e com outros algoritmos de agrupamento de comunidades para ver se as divisões encontradas pelo método de renormalização ficam mais coerentes com o que seria intuitivamente esperado.

BIBLIOGRAFIA

[1] Holme, P., **Metabolic robustness and network modularity: a model**, PloS one **6**, 2 (2011).

[2] Newman, M. E. J., **Networks: An Introduction**, Oxford University Press, (2010).

[3] Watts, D.J. e Strogatz, S.H., **Collective dynamics of "small-world" networks**, Nature **393**, 440 (1998).

[4] Albert, R. e Barabási, A.-L., **Emerging of Scaling in Random Networks**, Science **286**, 509 (1999).

[5] Wilson, K.G**. The renormalization group and critical phenomena**, Rev. Mod. Phys **55**, 583 (1983).

[6] Kadanoff, L. **Statistical Physics: Statics, Dynamics and Renormalization**, World Scientific (2000).

[7] García-Pérez, G., **Mercator: uncovering faithful hyperbolic embeddings of complex networks**, New Journal of Physics, **21**, 12 (2019).

[8] Rossi, R. A. e Nesreen K. Ahmed, N. K., **The Network Data Repository with Interactive Graph Analytics and Visualization,** Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (2015).

[9] García-Pérez, G. et al., **Multiscale unfolding of real networks by geometric renormalization**, Nature Physics **14**, 583 (2018).

[10] Zheng, M. et al. **Geometric renormalization unravels self-similarity of the multiscale human connectome**, Proceedings of the National Academy of Sciences, **117**, 33 (2020)

[11] Piovesana, Alan A., **GitHub deste projeto de iniciação científica**, acessado em agosto/2021. *https://github.com/Alandroid/network-renormalization*

[12] Blondel, Vincent D. et al., **Fast unfolding of communities in large networks**, Journal of Statistical Mechanics: Theory and Experiment, **2008**, 10 (2008).

[13] Newman, M. E. J., **Modularity and community structure in networks**, Proceedings of the National Academy of Sciences, **103**, 23 (2006).

[14] Traag, V. A. et al., **From Louvain to Leiden: guaranteeing well-connected communities**, Scientific Reports, **9** (2019).

[15] Csardi, G. e Nepusz, T., **The igraph software package for complex network research**, InterJournal, **Complex Systems** (2006).

[16] Aric A. Hagberg et al., **Exploring Network Structure, Dynamics, and Function using NetworkX**, Proceedings of the 7th Python in Science Conference (2008).

[17] Zachary, W. W., **An Information Flow Model for Conflict and Fission in Small Groups**, Journal of Anthropological Research, **33** (1977).