

ANÁLISE DAS REDES SOCIAIS À LUZ DA TEORIA DE GRAFOS DOS ESTUDANTES DE UMA UNIVERSIDADE PÚBLICA

Apiano Ferreira Morais¹, Daniela Leite Sousa², Hercules Souza Santana², Bethiele Milagre Leite²

Resumo

Neste artigo, realizamos uma análise das redes de “amizade” de 64 estudantes da Universidade Regional do Cariri através da rede virtual Facebook. Mostramos, através da caracterização das redes quanto ao grau de conectividade, coeficiente de modularidade, coeficiente de agregação e menor caminho entre os nós, que o comportamento social estereotipado de estudantes universitários de algumas áreas do conhecimento é confirmado estatisticamente pelos dados obtidos. Encontramos que uma rede inteiramente conectada formada pelas 64 redes do estudo resulta em uma rede de aproximadamente 29 mil nós com menor caminho médio 3,978. Esta rede tem um comportamento de decaimento exponencial em sua conectividade de aproximadamente 2,3 e características de redes de pequeno mundo. Ainda, encontramos que estatisticamente, homens e mulheres têm aproximadamente 30% a mais de amigos do sexo oposto em suas redes.

Palavras-chave: Redes complexas. Facebook. Redes Sociais.

ANALYSIS OF THE SOCIAL NETWORKS OF STUDENTS AT A PUBLIC UNIVERSITY THROUGH GRAPH THEORY

Abstract

In this paper we perform an analysis of networks of "friendship" of 64 students from the Regional University of Cariri through the Facebook virtual network. We show that the stereotyped social behavior of university students in some areas of knowledge is statistically confirmed by the data obtained, through the characterization of the networks in the degree of connectivity, modularity coefficient, clustering coefficient and shortest path between nodes. For a fully connected network built up by the 64 networks of the present study resulting in a network of approximately 29,000 nodes we found a minimal path of 3,978. This network has an exponential decay behavior in the connectivity about 2.3 and it exhibits characteristics of small-world network. Furthermore, we found that statistically, men and women have about 30% more friends of the opposite sex in their networks.

Keywords: Complex networks. Facebook. Social networks.

¹ Graduação em Física; Mestrado e doutorado em Física pela Universidade Federal do Ceará.. Atualmente é professor Assistente Doutor da Universidade Regional do Cariri;

² Acadêmico(a) da Universidade Regional do Cariri – URCA, Ceará - Brasil.

Introdução

A caracterização e o estudo de redes complexas teve grande impacto na ciência no início da década passada. Uma rede pode ser pensada como um conjunto de componentes, onde pares são conectados entre si através de ligações, o objeto de estudo está em toda parte: sistemas de transporte, teias alimentares, parceiros sexuais, relação entre enzimas e proteínas, rede dos agentes em um mercado, etc. Por esta razão, redes complexas reais têm sido tema de vários estudos científicos no tocante a caracterização e análise da dinâmica de suas propriedades topológicas: rede de comunicação via telefonia celular (ONNELA et al., 2007), redes de tráfego (GONZALEZ; HIDALGO; BARABASI, 2008), redes sociais e biológicas (AMARAL et al., 2000).

Dentre as redes sociais, uma ganha especial destaque devido ao seu grande número de componentes. The Facebook™ é uma marca de uma empresa estadunidense que proporciona através de uma página da internet um serviço de rede social. A página do Facebook foi inicialmente idealizada para conectar, através da internet, estudantes da Universidade de Harvard. Academicamente, as redes de amizade do Facebook foram estudadas por Rieder (2013) no contexto de análise de postagens e fração de membros da rede comentando o tópico específico. Lewis et al. (2008) estudaram as conexões levando em conta a distribuição étnica-racial e aspectos socioeconômicos de 1640 membros do Facebook. Em 2012, a equipe de Lars Backstrom mostrou que a rede inteira do Facebook (721 milhões de usuários ativos) tem menor caminho médio de 4,74, mostrando que o mundo virtual é bem menor do que se divulga (BACKSTROM et al., 2012).

Neste trabalho, desejamos responder perguntas do tipo: como se caracterizam as redes de amigos do Facebook de estudantes de uma instituição em particular? São estas redes livre-de-escala, aleatórias ou pequeno-mundo? Existem diferenças nas propriedades das redes de acordo com o gênero ou tipo de área de estudo dos estudantes da universidade?

Metodologia

Neste ponto, é conveniente fazermos uma distinção entre redes direcionadas e não direcionadas. A Figura 1-a mostra uma rede cujas ligações são de “mão-única”, ou seja, o nó A está ligado com o nó C, mas C não está ligado com o nó A. Dizemos que a direção da ligação é de A para C. Esta rede é então direcionada. Exemplos de redes direcionadas podem ser encontrados nas cadeias alimentares de um ecossistema, onde a predação é uma relação direcional. De maneira semelhante, as redes de transporte em uma cidade podem ser pensadas direcionadas, visto que cada rua pode ter apenas uma direção de fluxo.

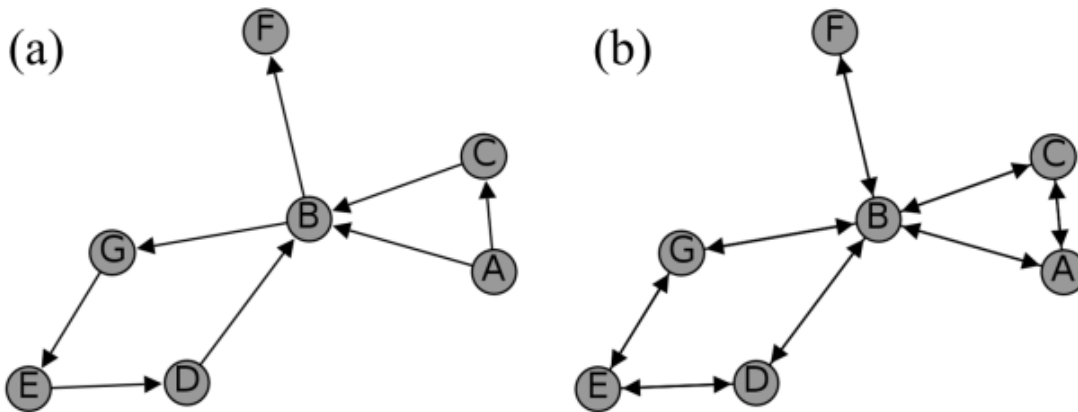


Figura 1 - Diagrama esquemático representando dois tipos de rede: (a) rede direcionada e (b) rede não-direcionada. As setas indicam o direcionamento das ligações em cada par de nós.

Já a Figura 1-b mostra uma rede onde todas as ligações são de “mão-dupla”. Este tipo de rede, chamada de não direcionada é encontrada em exemplos como as relações de amizade entre pessoas, relações de fronteira entre localidades e, em geral, nas conexões da malha ferroviária de um país.

Neste artigo, lidamos com os dois tipos de redes. A rede de amizades entre os indivíduos no Facebook é inerentemente não direcionada, visto que na rede de um indivíduo “A”, este tem uma relação de “amizade” com um indivíduo “B”, está relacionado de maneira recíproca para a rede de “B”. Definimos a relação de “amizade” como aquela em que o indivíduo “B” está classificado como “amigo” do indivíduo “A”. Nesta rede, cada indivíduo é um nó e a relação de amizade entre eles é uma ligação.

Obtenção das Redes

A metodologia empregada para a obtenção dos dados foi através de aquisição das conexões na rede de estudantes da universidade mediante ciência e acordo de sigilo de dados individuais. A aquisição se deu através do uso do aplicativo Netvizz (RIEDER 2013), que exporta estes dados para o formato legível pelo programa Gephi (BASTIAN 2009). Os dados obtidos são dispostos através de um arquivo de dados separados por vírgula contendo os nós (identificação do nome de usuário do amigo) e os pares de ligações entre os nós. Uma vez que os dados obtidos do Netvizz não integram o indivíduo na sua própria rede, um programa de reconfiguração de dados, escrito em linguagem Fortran, foi utilizado para incluir este indivíduo através da escrita da matriz de adjacência, A . Caso haja uma ligação entre os nós i e j , o elemento da matriz adjacência correspondente é $A_{ij} = 1$. Caso não haja ligação entre os nós, isto é, eles não se conhecem virtualmente, $A_{ij} = 0$. No caso de redes não direcionadas, $A_{ij} = A_{ji}$.

As análises das redes foram realizadas utilizando as rotinas do Gephi para obtenção da conectividade média, coeficiente de agregação médio, menor caminho médio, número de comunidades e coeficiente de modularidade.

Propriedades das redes

Utilizando conceitos obtidos da Teoria de Grafos, definimos a **conectividade** k_i de um nó i , como sendo o número de ligações que este apresenta. Através da matriz de adjacência, podemos escrever, para redes sem peso

$$k_i \equiv \sum_{j=1}^N A_{ij}, \quad 01$$

A **conectividade média de uma rede** com N nós é definida como a média simples da conectividade dos sítios:

$$\bar{k} \equiv \frac{1}{N} \sum_{i=1}^N k_i, \quad 02$$

Uma rede livre-de-escala é um grafo onde a distribuição de conectividade obedece a uma lei-de-potência. A fração de nós $P(k_i)$ na rede que apresentam conectividade k_i , para grandes redes, varia de acordo com

$$P(k_i) \propto k_i^{-\gamma}, \quad 03$$

Onde:

γ é um parâmetro a ser medido.

O menor caminho, ℓ_{ij} , entre dois nós i e j de uma rede é a quantidade de ligações mínimas necessárias para conectar estes dois nós. Em redes direcionadas, nem sempre é possível se chegar a todos os pontos, mas em redes não direcionadas, todos os pontos estão conectados. O menor caminho médio de um nó i na rede é a média dos menores caminhos entre este ponto e todos os outros na rede, $l_i \equiv \sum_{j=1}^N \ell_{ij} / N$. O **menor caminho médio da rede** é a média simples de todos l_i da rede:

$$\bar{l} \equiv \frac{1}{N} \sum_{i=1}^N l_i, \quad 04$$

Este parâmetro é capaz de classificar a rede através das propriedades de transporte de informação no grafo: se \bar{l} cresce com $\log(N)$, dizemos que a rede é de pequeno-mundo, em comparação com o conhecido o fenômeno social de **pequeno mundo** (SCHNETTLER, 2009). Redes sociais, internet e redes de genes todas exibem características de redes de pequeno mundo. Caso \bar{l} cresça com $\log[\log(N)]$, dizemos que a rede é ultra pequeno-

mundo. Cohen e Havlin mostraram analiticamente que redes livres-de-escala ($2 < \gamma < 3$) são deste último tipo (COHEN; HAVLIN, 2003).

O coeficiente de agregação local quantifica quão conectados são os vizinhos de um nó na rede (SOFFER; VASQUEZ, 2005). O coeficiente de agregação local C_i é dado pela proporção de ligações entre os nós dentro de sua vizinhança dividida pelo número de ligações que poderiam possivelmente existir entre eles. Para uma rede não direcionada, o nó i poderia ter $k_i(k_i - 1)/2$ ligações entre os nós de sua vizinhança. Então o **coeficiente de agregação local para uma rede não direcionada** pode ser escrito como

$$C_i = \frac{2n_i}{k_i(k_i - 1)}. \quad 05$$

Onde:

n_i é o número de nós conectados na vizinhança de i . O **coeficiente médio de agregação da rede** é definido como

$$\bar{C} \equiv \frac{1}{N} \sum_{i=1}^N C_i. \quad 06$$

Outro parâmetro fundamental na análise de grafos é a o **coeficiente, ou classe, de modularidade** de uma rede (NEWMAN, 2006). Seja S igual a 1, se o nó i pertence ao grupo r e igual a zero, caso contrário:

$$\delta_{ij} = \sum_r S_{ir} S_{jr} \quad 07$$

tal que,

$$Q \equiv \frac{1}{2m} \sum_{ij} \sum_r \left[A_{ij} - \frac{k_i k_j}{2m} \right] S_{ir} S_{jr}. \quad 08$$

Onde:

$2m = \sum_{i=1}^N k_i$ é o número de ligações na rede.

Redes com altos valores da classe de modularidade apresentam conexões robustas entre os nós dentro dos módulos, porém conexões esparsas entre nós em diferentes módulos. Em outras palavras, ela mede a robustez de como os grupos em uma rede estão conectados. Considerando que as redes estudadas são pequenas, em geral $N < 2^{12}$, o limite de resolução é considerado satisfatório (KUMPULA et al., 2007).

Resultados e discussão

Foram utilizados 64 grafos de amigos de estudantes distribuídos entre 14 cursos universitários contemplando 7 áreas do conhecimento diferentes: Ciências Exatas e Engenharias, Ciências da Saúde, Ciências

Humanas, Ciências Sociais e Artes. Em ordem a analisar o comportamento dos grafos em relação as áreas do conhecimento, agrupamos as áreas da seguinte maneira: artes (ART) engloba os cursos dos departamentos de Teatro e Artes Visuais; Ciências Biológicas e da Saúde (CBS) engloba os cursos dos departamentos de Biologia, Enfermagem e Educação Física; Ciências Exatas (CE) engloba os cursos dos departamentos de Física e Matemática; Ciências Humanas (CH) engloba os cursos dos departamentos de Geociências, História e Educação; Ciências Sociais (CS) engloba os cursos dos departamentos de Direito, Economia e Ciências Sociais; Engenharias (ENG) engloba os cursos dos departamentos de Construção Civil e Engenharia de Produção.

A Tabela 1 mostra os resultados obtidos segregados por grandes áreas do conhecimento, de onde é possível notar a pequena variação na média dos coeficientes de agregação e menor caminho médio das redes.

Tabela 2- Médias amostrais do coeficiente de modularidade, menor caminho médio, coeficiente de agregação e conectividade média para diferentes áreas do conhecimento.

Área	$\langle Q \rangle$	$\langle l \rangle$	$\langle C \rangle$	$\langle k \rangle$
ART	0,28565±0,6789	1,8963±0,0546	0,69666±0,12366	56,7710±18,9451
CBS	0,45911±0,8446	1,9098±0,03123	0,57878±0,04837	47,1234±9,1869
CE	0,33669±0,1453	1,8570±0,07935	0,63801±0,09186	60,1210±25,0597
CH	0,34267±0,1679	1,8917±0,05328	0,57222±0,04864	68,5952±21,0586
CS	0,52371±0,0315	1,9380±0,01964	0,54857±0,03425	42,1397±2,0801
ENG	0,36465±0,1236	1,87972±0,0542	0,36465±0,11661	80,2008±38,9123

A média amostral da conectividade média das redes foi 64.5774 com desvio padrão de 21.5674. Na figura 2 é mostrado como a conectividade média varia de acordo com a área do conhecimento dos cursos estudados. Observamos que a conectividade das redes dos estudantes de ciências sociais foi a menor encontrada (42,1397±2,0801). As outras conectividades são bastante aproximadas se considerarmos a barra de desvio.

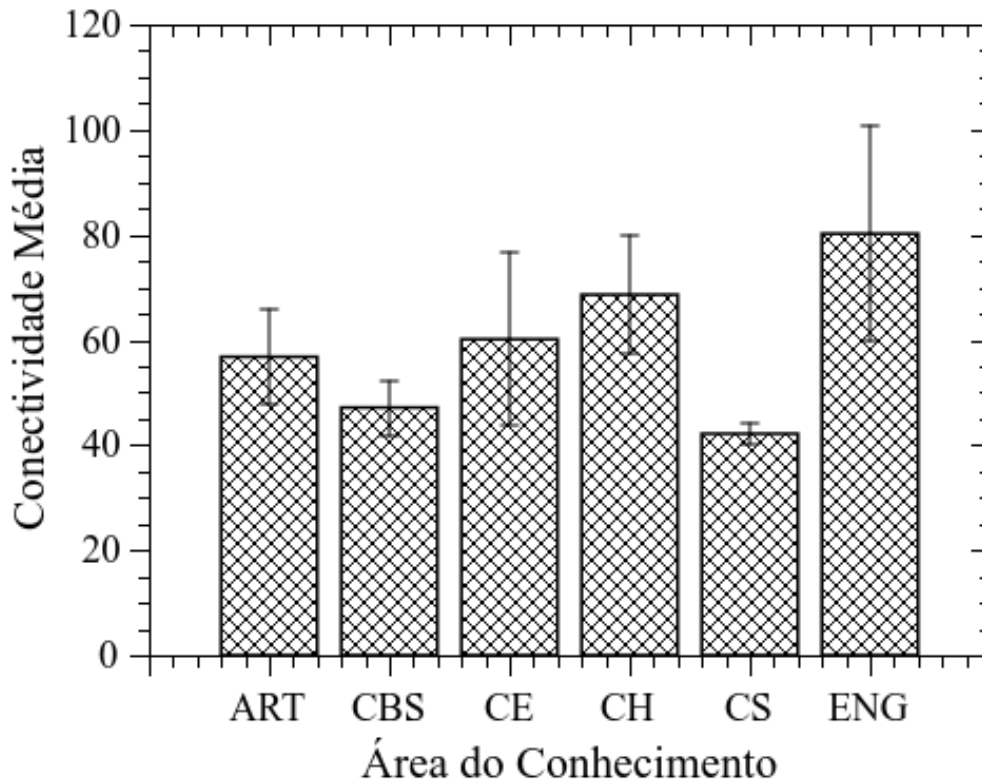


Figura 2 - Conectividade media dos grafos separados por área do conhecimento.

Observando também a Figura 3, que mostra o coeficiente de modularidade médio das redes por área, encontramos que a área com maior conectividade média é justamente a área das Ciências Sociais ($0,52371 \pm 0,0315$). Isto mostra que embora os estudantes desta área sejam mais sociáveis, uma vez que apresentam mais módulos distintos nas redes, eles apresentam poucos amigos. Em geral, isto acontece com indivíduos que participam de diversos grupos sociais. Este comportamento é partilhado, com menos intensidade, pelos estudantes da área de Ciências Biológicas e da Saúde. Os resultados mostram também que os estudantes da área de Artes e Engenharia são socialmente mais fechados e tem relações de amizades no Facebook, em geral, com membros de um mesmo grupo (Figura 3).

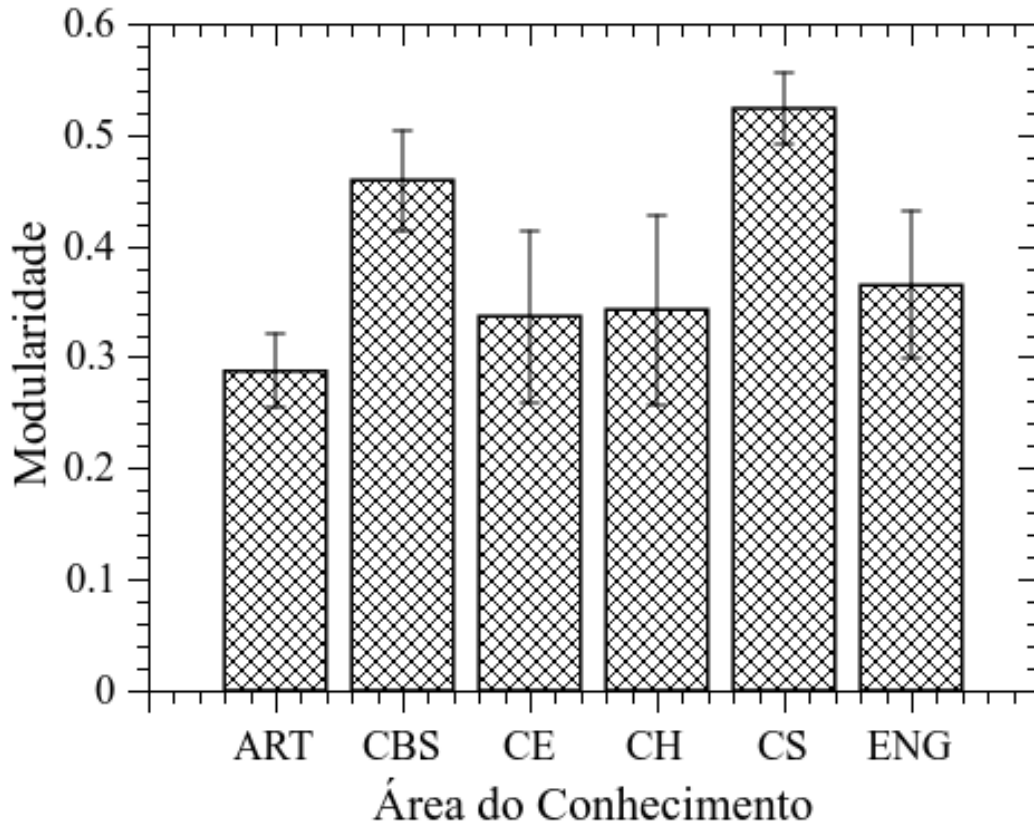


Figura 3—Média do coeficiente de modularidade para diversas áreas do conhecimento.

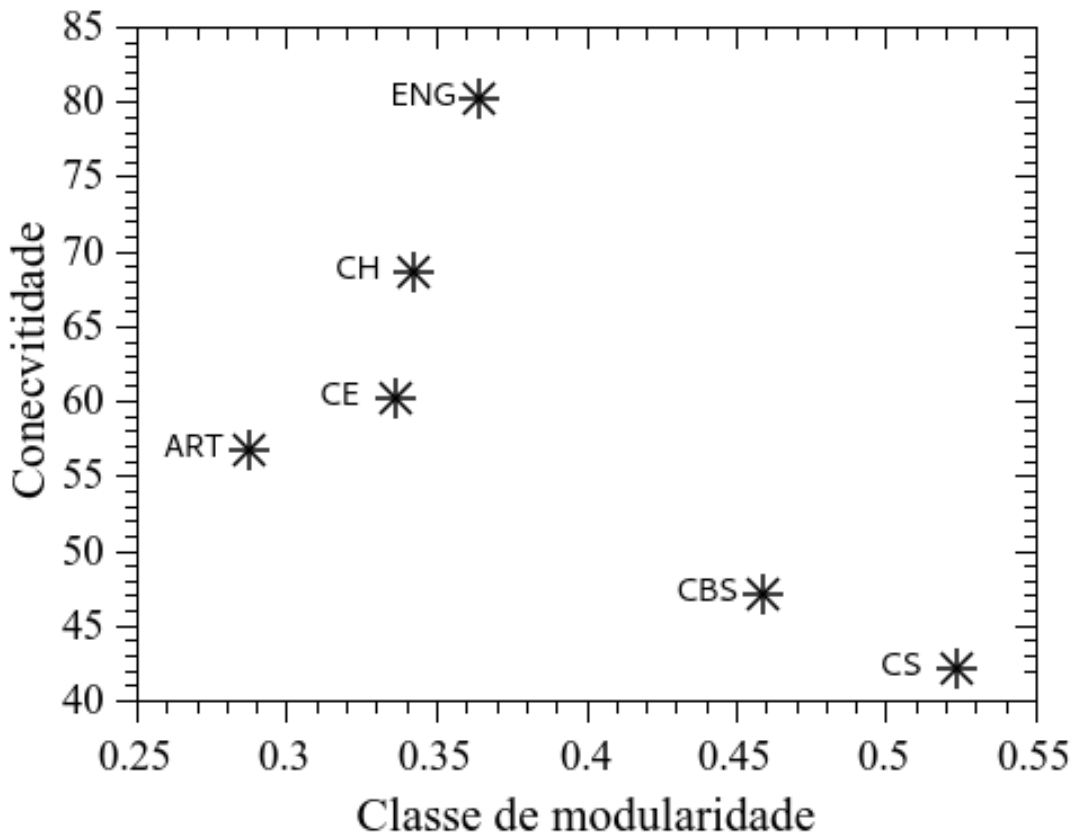


Figura 4 - Espaço de fase da média da classe de modularidade versus média das conectividades das redes separadas por área.

Em relação à classe de modularidade, entre os grupos de homens ($0,39857 \pm 0,0651$) e mulheres ($0,37752 \pm 0,0717$), não obtivemos diferenças significativas. O mesmo ocorre com a média da conectividade, média do coeficiente de agregação e tamanho médio das redes (Tabela 2). Porém, um dado interessante emerge quando analisamos a razão de indivíduos que declaram seu gênero. A razão aqui é definida como o número de mulheres (N_M) na rede dividido pelo número de homens (N_H). Os dados de redes de estudantes do sexo feminino mostram que a média desta razão é 1,2437 com desvio 0,1539 e para os homens é 0,8746 com desvio de **0,1472**.

Tabela 3- Dados agrupados do coeficiente de modularidade, tamanho médio da rede, média da conectividade e média da razão entre homens e mulheres das redes de homens e mulheres.

Sexo	$\langle Q \rangle$	$\langle N \rangle$	$\langle k \rangle$	$\langle N_M/N_H \rangle$
Masculino	$0,39857 \pm 0,0651$	$641,167 \pm 41,758$	$57,056 \pm 14,251$	$1,2437 \pm 0,1539$
Feminino	$0,37752 \pm 0,0717$	$786,032 \pm 80,449$	$70,250 \pm 26,923$	$0,8746 \pm 0,1472$

Por último, em ordem a ter uma estatística robusta, integramos todas as amostras de redes estudadas em uma “super-rede”. A Figura 5 mostra a representação desta rede, evidenciando diversas comunidades conectadas entre si, em geral, por mais de um nó. Aparentemente, esta rede é robusta e pode ser utilizada para propagar informações eficientemente. A Figura 6 mostra o histograma da conectividade da super-rede, que indica claramente se tratar de uma rede de pequeno mundo com decaimento exponencial (ALBERT et al., 2000). De fato, a figura 6 mostra dois tipos de comportamento de decaimento exponencial. Para pequenos valores da conectividade, o decaimento é mais acentuado (expoente $8,375 \pm 0,026$) e, para valores grandes da conectividade, o decaimento é mais suave (expoente $2,292 \pm 0,039$). Este comportamento de decaimento na cauda da distribuição é encontrado em outras redes sociais e em redes neuronais do verme *C. elegans* (AMARAL et al., 2000).

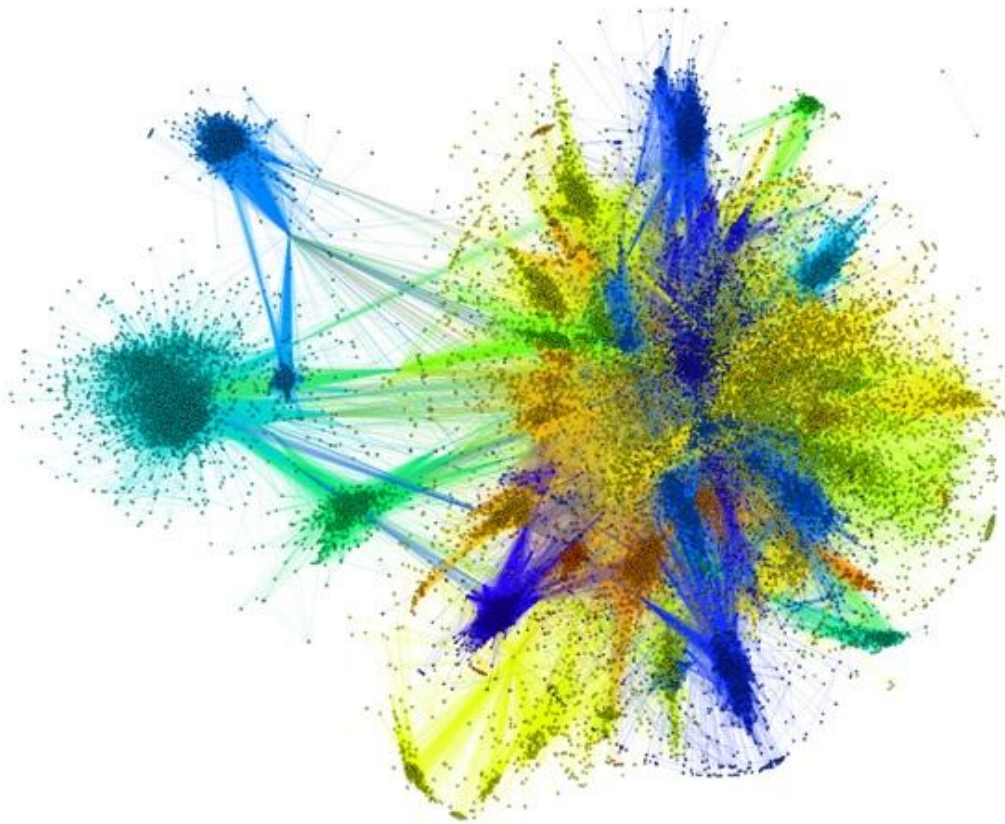


Figura 5 - Representação esquemática tipo Yifan-Hu da rede integrada de todas as amostras estudadas ($N = 28912$) com as ligações entre os nós coloridas pela classe de modularidade.

O menor caminho médio para esta rede é 3,978. Este resultado está longe dos seis graus de separação proferidos pelo psicólogo Stanley Milgram (KLEINFELD 2002), mas se aproxima bastante do menor caminho médio para a relação entre atores em filmes (WATTS; STROGATZ 1998). Acreditamos que o número de nós na super-rede ainda seja insuficiente para se chegar ao resultado conjecturado por Milgram (caso ele esteja correto), uma vez que em redes de pequeno mundo \bar{l} cresce com $\log N$ e teríamos de ter uma rede com mais de cinco milhões de nós!

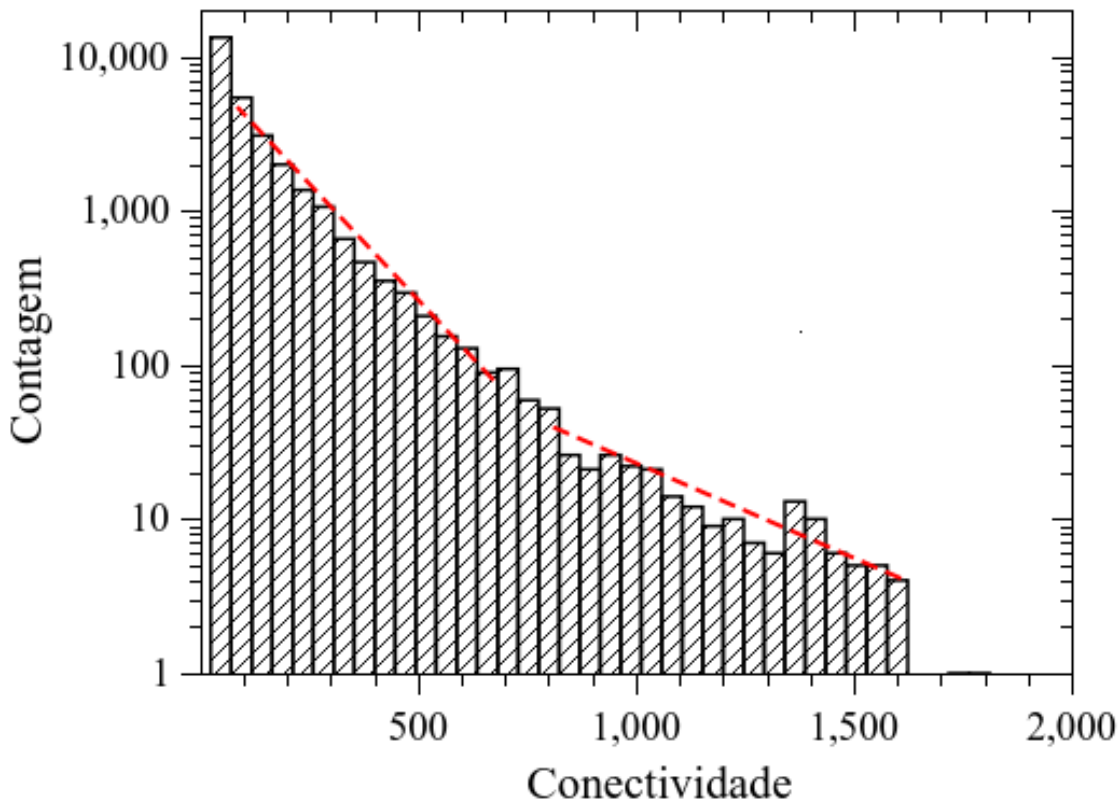


Figura 6 - Histograma da log-normal da conectividade da rede integrada de todas as amostras obtidas no estudo. As linhas pontilhadas são curvas de ajuste de decaimento exponencial. Para baixos valores da conectividade o expoente encontrado foi $8,375 \pm 0,026$ e para altos valores foi $2,292 \pm 0,039$.

Conclusões

Apesar da quantidade pequena de dados, foi possível mostrar tendências de comportamento e caracterização das redes sociais dos estudantes da universidade estudada com base em suas conexões. Uma vez que as características das redes não variam significativamente com o sexo dos entrevistados, podemos afirmar que o comportamento de encerramento em poucos círculos de amizade encontrado nas áreas de Artes e Engenharia, em contraste com a área de Ciências Sociais, é um comportamento característico daquelas áreas. Esta quantificação confirma estatisticamente o estereótipo de que estudantes de cursos da área de Ciências Sociais (Ciências Sociais, Direito e Economia) são mais abertos em seu comportamento social, ao passo que estudantes da área de Artes e Engenharias têm uma tendência a não se relacionar com outras áreas.

Outro resultado interessante é que homens, em geral, têm mais amigos do sexo feminino e mulheres tem mais amigos do sexo masculino. Este comportamento nunca tinha sido reportado e ajuda a mostrar que aparentemente a razão da criação do Facebook ainda vigora sobre suas outras utilizações.

A rede social gerada pela rede de todos os entrevistados deu uma pista de que as redes do Facebook são realmente de pequeno mundo semelhante a outras redes sociais. Este talvez seja o ponto chave deste negócio comercial, uma vez que é sabido que redes de pequeno mundo são mais eficientes no transporte de informação

(LATORA; MARCHIORI, 2001), isto é, a rede complexa mais hábil quando se deseja espalhar conhecimento; seja ela uma notícia, fofoca ou propaganda. Entretanto, o comportamento de seis graus de separação proposto pelo psicólogo Stanley Milgram não é corroborado pelos nossos dados. De fato, este conceito tem caído em descrédito na última década com experimentos capazes de detectar os menores caminhos em várias redes sociais.

Agradecimentos

Os autores agradecem a Universidade Regional do Cariri e à Fundação Cearense de Apoio à Pesquisa (FUNCAP) pelo apoio financeiro através de bolsas do PIBIC.

Referências

- ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Error and attack tolerance of complex networks. **Nature**, v. 406, n. 6794, p. 378-382, 2000.
- AMARAL, L.A.N. et al. Classes of small-world networks. **Proceedings of the national academy of sciences**, v. 97, n. 21, p. 11149-11152, 2000.
- BACKSTROM, L. et al. Four degrees of separation. In: **Proceedings of the 4th Annual ACM Web Science Conference**. ACM, 2012. p. 33-42.
- BASTIAN, M. et al. Gephi: an open source software for exploring and manipulating networks. **ICWSM**, v. 8, p. 361-362, 2009.
- BLONDEL, V.D. et al. Fast unfolding of communities in large networks. **Journal of Statistical Mechanics: Theory and Experiment**, v. 2008, n. 10, p. P10008, 2008.
- COHEN, R.; HAVLIN, S. Scale-free networks are ultrasmall. **Physical review letters**, v. 90, n. 5, p. 058-701, 2003.
- GONZALEZ, M.C.; HIDALGO, C.A.; BARABASI, A.-L. Understanding individual human mobility patterns. **Nature**, v. 453, n. 7196, p. 779-782, 2008.
- KLEINFELD, J. Could it be a big world after all? The six degrees of separation myth. **Society**, April, v. 12, p. 5-2, 2002.
- KUMPULA, J.M. et al. Limited resolution in complex network community detection with Potts model approach. **The European Physical Journal B-Condensed Matter and Complex Systems**, v. 56, n. 1, p. 41-45, 2007.
- LATAPY, M. Main-memory triangle computations for very large (sparse (power-law)) graphs. **Theoretical Computer Science**, v. 407, n. 1, p. 458-473, 2008.
- LATORA, V.; MARCHIORI, M. Efficient behavior of small-world networks. **Physical review letters**, v. 87, n. 19, p. 198701, 2001.
- LEWIS, K. et al. Tastes, ties, and time: A new social network dataset using Facebook. com. **Social networks**, v. 30, n. 4, p. 330-342, 2008.

ONNELA, J.-P. et al. Structure and tie strengths in mobile communication networks. **Proceedings of the National Academy of Sciences**, v. 104, n. 18, p. 7332-7336, 2007.

NEWMAN, M.E.J. Modularity and community structure in networks. **Proceedings of the National Academy of Sciences**, v. 103, n. 23, p. 8577-8582, 2006.

SOFFER, S.N.; VÁZQUEZ, A. Network clustering coefficient without degree-correlation biases. **Physical Review E**, v. 71, n. 5, p. 057101, 2005.

SCHNETTLER, S. A structured overview of 50 years of small-world research. **Social Networks**, v. 31, n. 3, p. 165-178, 2009.

RIEDER, B. Studying Facebook via data extraction: the Netvizz application. In: **Proceedings of the 5th Annual ACM Web Science Conference**. ACM, 2013. p. 346-355.

WATTS, D.J.; STROGATZ, S.H. Collective dynamics of 'small-world' networks. **Nature**, v. 393, n. 6684, p. 440-442, 1998.



Recebido: 19/05/2015

Aceito: 25/07/2015