

Métricas de Centralidade em Redes Sociais

Paula Alexandra Laranjeira

Licenciada Informática, Univ. Aberta, xana_gouveia@sapo.pt

Luís Cavique

Univ. Aberta, lcavique@uab.pt

Resumo

Nos modelos de redes sociais, tal como na teoria de grafos, os vértices representam os atores e as arestas ou arcos a relação entre eles. Atores influentes são aqueles que estão frequentemente envolvidos na relação com outros atores. Este envolvimento torna-os mais visíveis sendo considerados mais centrais na rede. É neste sentido que as métricas de centralidade tentam descrever as propriedades da localização de um nó fulcral numa rede. Estas medidas têm em consideração os diferentes modos de interação e comunicação de um ator com os restantes elementos, sendo mais importantes, ou centrais, aqueles que estão localizados em posições mais estratégicas na rede. Neste trabalho apresenta-se o estudo de cinco métricas de centralidade: grau, proximidade, intermediação, vetor próprio e katz. Descrevem-se os algoritmos implementados no cálculo das medidas e apresenta-se um caso de estudo. Para completar o estudo é apresentada uma análise comparativa entre os resultados obtidos no aplicativo NodeXL, e os resultados obtidos através dos algoritmos implementados.

palavras-chave: redes sociais, centralidade, proximidade, intermediação, vetor próprio, Katz

Abstract

Considering models for social networks as graphs, nodes represent the actors and the edges represent the relationship between them. Influential actors are the ones that are frequently involved on relationships between other actors. This involvement makes them more visible and considered more central on the network. In this sense centrality metrics try to describe the localization properties of an important node of the network. These measures have in consideration the different interaction and communication modes an actor has with others, being more important or central the ones that are located on more strategic locations on the network. On this work it is presented the study of five centrality measures: degree, closeness, betweenness, eigenvector and katz. It is made a description of the algorithms implemented, and it is presented a case study. To complete the study it is also made a comparative analysis between results obtained with NodeXL, and the results from the algorithms implemented.

keywords: social network, centrality, closeness, betweenness, eigenvector, katz

1. Introdução

Segundo alguns investigadores os seres humanos têm muito mais predisposição para reconhecer padrões do que para pensar através de combinações lógicas, pois segundo estes, este tipo de reconhecimento está implícito e é mais facilmente assimilado pelo circuito neurológico humano.

Segundo Steven Johnson [Johnson 2001], essas faculdades compensam a velocidade extremamente baixa dos neurónios, que são milhões de vezes mais lentos que alguns computadores a resolver problemas. No entanto, por ser um sistema paralelo de grande porte, com 100 milhares de milhões de neurónios a trabalhar em simultâneo, o cérebro pode protagonizar proezas admiráveis no que concerne ao reconhecimento de padrões, como a identificação ou interpretação de imagens. Com esta capacidade, as pessoas podem classificar situações, vivências, experiências, formando uma imagem bem definida de dados muitas vezes difíceis de serem interpretados pelo computador. Assim, se o mapeamento estrutural gráfico das redes sociais pudesse ser disponibilizado, o cérebro humano utilizaria a sua principal habilidade cognitiva para prover informações cruciais, mas despercebidas às melhores máquinas.

O propósito do estudo de métricas de redes é permitir entender tanto a estrutura dos sistemas conectados quanto as formas pelas quais diferentes tipos de influência se propagam através deles. Por padrão entende-se uma figuração específica que se caracteriza por uma regularidade, repetição de partes e acumulação de elementos. Obviamente, o reconhecimento de padrões e a compreensão das suas origens constituirão apenas um primeiro passo. Será também necessário saber como interagir e como utilizar as propriedades das redes a favor da causa humana [Watts 2003].

Na análise clássica de redes sociais existem dois conjuntos de métricas: a centralidade dos nós e as componentes conexas [Scoot 2000]. Neste trabalho iremos discutir e comparar diferentes formas de calcular a centralidade. Ao conceito de centralidade associa-se o conceito de poder e de influência dos atores adjacentes.

Este artigo está estruturado em 7 secções. Nas Secções 2 e 3 são definidos conceitos básicos de redes sociais e grafos. Na Secção 4 são apresentadas as cinco métricas de centralidade das redes sociais que iremos desenvolver neste artigo. Na Secção 5 são apresentados os algoritmos desenvolvidos e a sua implementação. Na Secção 6 os resultados computacionais são detalhados. Finalmente da Secção 7 são elaboradas as conclusões.

2. Redes Sociais

Uma rede pode ser expressa matematicamente por um grafo que é constituído por um conjunto de pontos, os nós ou vértices, conectados por linhas que expressam uma relação entre eles, as arestas. O conceito de Rede Social pode ser descrito como uma estrutura social composta de nós que estão relacionados por um ou vários tipos de relações. Os nós representam pessoas ou organizações, e as relações podem ser de comunicação, financeiras, profissionais, amorosas, etc.

Reduzida ao esqueleto básico, uma rede constitui simplesmente um conjunto de objetos conectados entre si de certa forma. *Routers* na Internet, neurónios num cérebro, rotas de

aviação, etc. são exemplos de redes, embora distintas. Atualmente, o ramo da matemática conhecido como Teoria dos Grafos, proposto originalmente pelo matemático Leonhard Euler, constitui a base da modelação matemática para o estudo da ciência das redes.

A Teoria de Grafos tem sido muito utilizada nas análises de redes sociais devido à sua capacidade representacional e simplicidade. Basicamente, um grafo é constituído por nós (n) e ligações (l) ou arestas (E de *edge*) que conectam os nós. Em redes sociais a representação por grafos também é chamada de sociograma, em que os nós são os atores ou eventos e as linhas de ligação estabelecem o conjunto de relações num desenho bidimensional.

Uma rede social é um conjunto de pessoas ou grupos de pessoas, ligados entre si por relações que podem ser profissionais, familiares ou outras. Os estudos neste campo iniciaram-se nos anos trinta por Moreno [1934] com a representação gráfica das redes sociais, os sociogramas. A introdução de modelos matemáticos foi iniciada por Rapoport [1957] e continuada por Erdos e Renyi [1959] na criação dos grafos aleatórios. Outro estudo importante, na área da sociologia, foi desenvolvido por Milgram [1969] que levaram ao conceito de "pequeno mundo" (*small world*) e à noção dos "seis graus de separação". A experiência consistiu no envio de cartas entre dois pontos distintos, do Nebraska, no meio-oeste, até à costa leste dos EUA, em Massachusetts. Foi pedido às pessoas que não utilizassem o correio mas os seus contactos para enviarem as cartas. As cartas que chegaram ao seu destino precisaram de passar, em média, por cerca de 6 pessoas intermediárias, o que levou Milgram a concluir que, neste caso, os americanos estão aproximadamente à distância de 6 passos entre si.

O conceito de redes sociais ganhou uma nova dinâmica com a emergência da Web 2.0. Em Junho de 2008, o Facebook ultrapassou mundialmente o MySpace em número total de utilizadores, tornando-se a maior rede social online já existente. Ellison, Steinfeld, Lampe [2007] fizeram um estudo amplo indicando como o Facebook, enquanto suporte das relações sociais, modificou os processos sociais, permitindo que atores que não tinham mais contacto devido à distância pudessem investir ainda nesses laços sociais. Essas características mostram-nos que há uma mudança no suporte da interação, que vai permitir que as conversações, antes mapeadas unicamente pela observação, sejam facilmente registadas. Pela primeira vez, graças aos recursos proporcionados pela Web 2.0, como os sites de rede social, os atores passam a registar as suas conversas, as suas interações e as suas redes. E com isso, o mapeamento dessas redes ganha novo potencial no sentido de que, pela primeira vez, é possível mapear gostos, atos, ideias e conexões de milhares de pessoas, procurar e estabelecer padrões entre essas múltiplas redes, principalmente através das interações que são mediadas por essas ferramentas. O estudo das redes sociais é, portanto, o estudo dos padrões sociais. Com a Web esses padrões tornam-se mais evidentes, há a possibilidade de estudá-los de uma forma mais abrangente e em maior escala.

Depois das abordagens das redes sociais estudadas pela sociologia e pela matemática, emerge uma terceira escola, a das Redes Complexas. O trabalho pioneiro de Watts e Strogatz [1998] publicado na *Nature*, estuda grafos com propriedades de pequenos-mundos e com grau de distribuição com ausência de escala (*free-scale*). Segue-se o

modelo de Barabási e Albert [1999] e os modelos epidemiológicos SIR que vêm definir novas fronteiras para as redes em sistemas dinâmicos.

3. Grafos

3.1. Algumas propriedades

Os grafos são formas abstratas utilizadas na representação de redes. Um grafo G é definido pelo par ordenado (V,E) , em que V é um conjunto de vértices e E é um conjunto de arestas. Cada aresta e pertencente ao conjunto E denota-se por $e = (v,w)$ sendo este um par de vértices. Os vértices v e w são os extremos da aresta e são denominados vértices adjacentes ou vizinhos. A aresta e é dita incidente a ambos os vértices v e w , ou seja, a aresta parte de um e chega ao outro vértice.

O Grafo não direcionado é aquele em que a relação (v,w) é simétrica. Neste caso, existe uma aresta direcionada que une v e w , sendo que o contrário também ocorre. Em consequência disso, podemos dizer que as arestas que ligam os vértices não possuem orientação.

A Figura 1 representa um grafo formado pelo conjunto de vértices $V = \{a,b,c,d,e,f,g\}$, e cada par de vértices que se conectam formam as arestas, pertencentes ao conjunto $E = \{(a,c), (b,c), (c,d), (d,e), (d,f), (d,g)\}$.

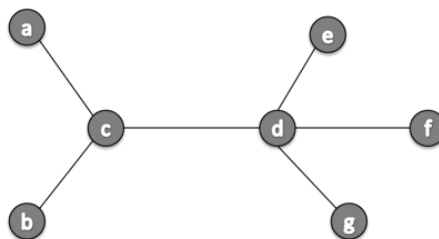


Figura 1 - Grafo não direcionado

Um grafo $G(V,E)$ é conexo quando existir um caminho entre cada par de vértices, caso contrário, é dito desconexo. Desse modo, se existir pelo menos um par de vértices que não esteja conectado através de um caminho o grafo é desconexo. Na Figura 2 temos um grafo conexo em (a), pois sempre podemos encontrar um caminho entre quaisquer pares de vértices escolhidos. Em (b) isso não acontece, pois se escolhermos o nó 3 e o nó 1, verificamos que não existe nenhum caminho conectando esse par de nós.



Figura 2 (a) - Grafo conexo

(b) Grafo desconexo

3.2. Representação da Matriz de adjacência

Diversas representações matriciais resultam de estruturas de grafos e o seu uso está, habitualmente, associado à necessidade da realização de cálculos envolvendo dados estruturais. A matriz mais utilizada é a matriz de adjacência (A), ver Figura 3. Esta é uma matriz quadrada, cuja ordem n corresponde ao número de vértices que o grafo possui; cada linha e cada coluna desta matriz correspondem a um vértice do grafo. Os elementos dessa matriz estão associados aos dados estruturais do grafo $G(V, E)$, ver Figura 4. Estes elementos possuem valores não nulos quando existem conexões entre os vértices correspondentes aos elementos e valores nulos caso contrário. Quando se tratam de grafos valorados, nos quais as arestas possuem pesos, os valores dos elementos não nulos da matriz de adjacência recebem o peso da aresta. Para grafos não valorados, os elementos não nulos recebem o valor 1.

$$A = [a_{ij}] \quad a_{ij} = \begin{cases} 1, & \text{se } \exists (i, j) \in E \\ 0, & \text{se } \nexists (i, j) \in E \end{cases}$$

Figura 3 - Definição de matriz de adjacência de grafo não valorado



Figura 4 - Grafo e respetiva matriz de adjacência

A matriz de adjacência de um grafo não direcionado é uma matriz simétrica.

4. Medidas de Centralidade

As medidas de centralidade surgem no contexto da análise de redes sociais [Landherr, Friedl, Heidemann 2010]. Intuitivamente, numa rede, os vértices mais centrais são aqueles a partir dos quais podemos atingir qualquer outro com mais facilidade ou rapidez. As medidas de centralidade identificam a posição de um indivíduo relativamente a outros na sua rede. A centralidade traduz também a ideia de poder. Assim, quanto mais central o indivíduo, maior influência e poder terá na sua rede [Gama 2012].

4.1. Centralidade de Grau

A conceção mais simples e intuitiva no que diz respeito à centralidade de um vértice é o número de contatos diretos que ele possui. Uma pessoa que se encontra numa posição que permite o contato direto com muitos outros é vista pelos demais como um canal maior de informação, razão pela qual dizemos ser mais central. Assim, a centralidade de grau nada mais é que a contagem do número de adjacências de um vértice.

Seja G um grafo qualquer (conexo ou não) com n vértices e seja x um vértice de G . A centralidade de grau de x denotada por $\sigma_D(x)$, é o número de arestas incidentes a x . Recorrendo à matriz de adjacências do grafo, tem-se:

$$\sigma_D(x) = \sum_{i=1}^n a_{ix} \quad (1)$$

onde a_{ix} são elementos da matriz de adjacência $A(G)$.

4.2. Centralidade de Proximidade

Uma outra medida [Sabidussi 1966] designa-se centralidade de proximidade e tem por base a soma das distâncias de um vértice em relação aos demais vértices do grafo. Em diferentes contextos, mais importante que ter muitas ligações é não estar longe demais dos restantes elementos.

Seja G um grafo conexo com n vértices e seja x um vértice de G . A centralidade de proximidade de x é dada pelo inverso da soma das distâncias de x a todos os outros vértices do grafo, ou seja,

$$\sigma_C(x) = \frac{1}{\sum_{i=1}^n d_G(x,i)} \quad (2)$$

onde $d_G(x, i)$ corresponde à distância de x a i .

4.3. Centralidade de Intermediação

Esta medida permite medir a capacidade que um ator tem de poder influenciar os seus pares numa rede. A ideia desta medida pode facilmente compreender-se na seguinte analogia: uma cidade integrada em várias rotas comerciais terá certamente uma vantagem estratégica, o que motiva a seguinte ideia: um nó importante faz parte de muitos caminhos.

Seja G um grafo (conexo ou não) com n vértices e seja x um vértice de G . A centralidade de intermediação de x é dada por:

$$\sigma_B(x) = \sum_{i=1}^n \sum_{i < j} \frac{g_{ij}(x)}{g_{ij}}, i, j \neq x \quad (3)$$

onde g_{ij} representa o número de caminhos mínimos do vértice i ao vértice j e $g_{ij}(x)$ indica a quantidade desses caminhos mínimos que passam por x .

4.4. Centralidade do Vetor Próprio

Esta medida tem por base os conceitos de valor próprio e vetor próprio da matriz de adjacência do grafo G . Refere-nos que outra maneira de interpretar a centralidade é pensar que a centralidade de um vértice está em função dos vértices vizinhos, ou seja, que um nó importante tem vizinhos importantes [Bonacich, Lloyd 2001].

Considere-se que a centralidade do vetor próprio, $\sigma_E(x)$, é proporcional (por um fator λ^{-1}) à soma das centralidades dos seus vizinhos e tem-se que:

$$\sigma_E(x) = \frac{1}{\lambda_{\max}(A)} \sum_{j=1}^n a_{jx} \cdot v_j \quad (4)$$

onde $\lambda_{\max}(A)$ é maior valor próprio (em módulo) da matriz de adjacências A e $v = (v_1, \dots, v_n)^T$ é o vetor próprio associado a ele.

A escolha do valor próprio de maior valor absoluto da matriz de adjacência [Bonacich 2001] tem por base o teorema de Perron-Frobenius:

Teorema Perron-Frobenius - Seja A uma matriz quadrada, irredutível, de ordem n com entradas não negativas. Então existe um vetor próprio v de componentes todas positivas e o correspondente valor próprio ρ é simples e tal que para qualquer valor próprio λ de A , $|\lambda| \leq \rho$.

4.5. Centralidade de Katz

Ao efetuar uma combinação linear do número total de caminhos de tamanho arbitrário de x para outros vértices, por exemplo, atribuindo um peso k^i aos caminhos de tamanho i , obtém-se a centralidade de Katz, proposta por este em 1953, na qual:

$$\sigma_K(x) = \vec{1}^T (\sum_{i=1}^{\infty} k^i A^i) \vec{e}_x \quad (5)$$

onde: $\vec{1} = (1,1, \dots, 1,1)^T$, $\vec{e}_x = (0, \dots, 0, 1, 0, \dots, 0)^T$ e matriz A^i cuja entrada a_{xy} representa o número de caminhos de comprimento i de x a y .

O teorema apresentado a seguir garante a convergência da série $\sum_{i=1}^{\infty} k^i A^i$ e deste forma permite simplificar a métrica de Katz [1953].

Teorema : *Seja λ_1 o maior valor próprio de A .*

$$\text{Se } |k| < \frac{1}{\lambda_1}, \text{ então } \left(\sum_{i=1}^{\infty} k^i A^i \right) \rightarrow (I_n - kA)^{-1} - I_n$$

A versão simplificada da Centralidade de Katz é obtida para um valor de k mais pequeno do que o inverso do máximo valor próprio de A (matriz adjacência do grafo G) tem-se:

$$\sigma_K(x) = \vec{1}^T ((I_n - kA)^{-1} - I_n) \vec{e}_x, \quad (6)$$

sendo I_n a matriz identidade de dimensão $n = |V_G|$.

5. Implementação das métricas de centralidade

O *software* que permite determinar as métricas referidas e cuja *interface* pode ser observada na Figura 5 foi construído recorrendo a duas linguagens de programação: o Java e o *software* de computação estatística R. Usou-se o R devido ao seu poderoso cálculo vetorial e matricial, tendo sido usado para implementar os *scripts* que calculam as métricas de grau, vetor próprio e katz e ainda desenhar o grafo e os gráficos de todas as medidas em estudo. O Java foi usado para implementar outros algoritmos, tais como as adaptações dos algoritmos de Floyd [1962] e de Dijkstra [1959], suportar todo o código base da aplicação, e ainda efetuar a ligação entre a GUI e os *scripts* de R. Apresenta-se de seguida uma síntese do trabalho realizado.



Figura 5 -Interface gráfica do software implementado

5.1. Implementação de scripts de R

As *scripts* que desenham o grafo e os gráficos de todas a métricas e que calculam as medidas de centralidade de Grau, Vetor Próprio e Katz, com base na matriz adjacência, foram implementados em *scripts* do R. Para as restantes duas medidas de centralidade, proximidade e intermediação, apenas se usou o R para elaborar os gráficos. Os algoritmos de encaminhamento fundamentais à sua determinação foram implementados em Java.

No menu Grafo foi implementada a *script* *grafo.R* que permite o desenho do grafo de uma determinada rede (vide Figura 6) e que tem por base a biblioteca *igraph* do R, implementando métodos que permitem desenhar o grafo e determinar a matriz adjacência. Todas as outras métricas dependem da correta execução desta. Estas *scripts* usam as potencialidades do R, nomeadamente o cálculo matricial.

Na implementação da *script* que determina a medida de centralidade de vetor próprio recorre-se a funções do R que permitem o cálculo do valor próprio de maior valor absoluto da matriz de adjacência e respetivo vetor próprio associado.

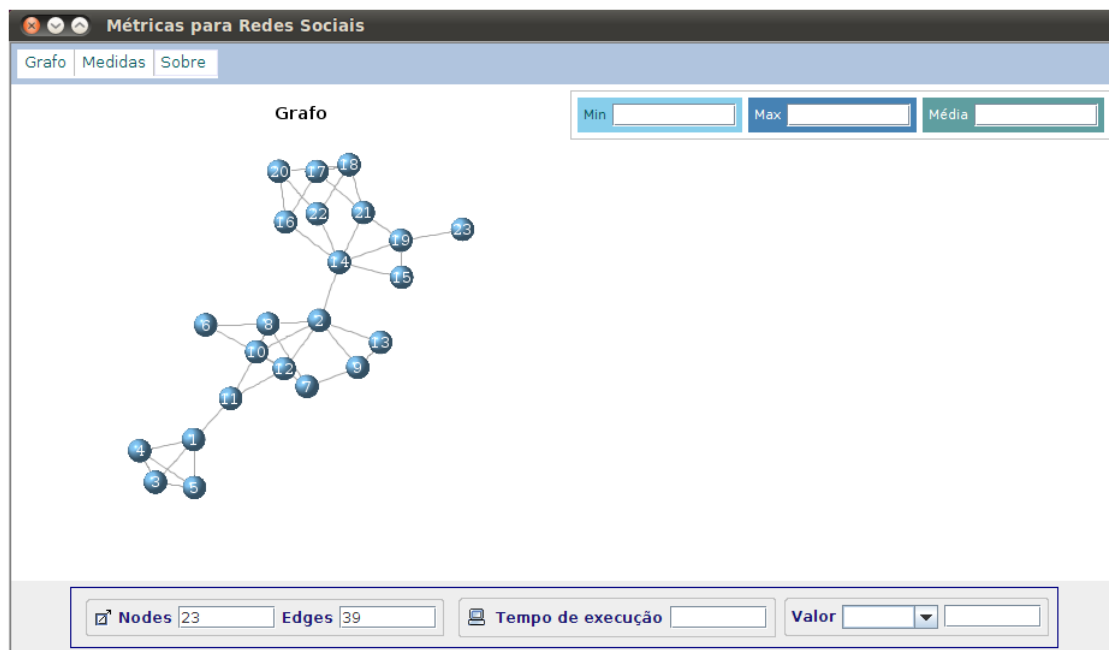


Figura 6 - Rede de estudo com 23 nós e 39 arestas

A invocação de funções do R a partir da aplicação Java foi implementada recorrendo à API JRI (agora integrada do projeto rJava). O processo de invocação consiste em instanciar um objeto da classe Rengine. Este objeto é depois usado para instanciar variáveis em R, passar instruções e obter o valor de variáveis. Para a integração de gráficos construídos em R em componentes Swing recorreu-se a outra API JavaGD.

5.2. Implementação em Java

Foram implementadas duas adaptações de dois importantes algoritmos teóricos das ciências das redes: Algoritmo de Floyd e Algoritmo de Dijkstra. E tal como pode observar-se na Figura 7, é dada a possibilidade de escolha ao utilizador sobre o algoritmo pretendido para o cálculo das métricas de intermediação e também da proximidade. Esta escolha serve apenas para enriquecer o *software* e possibilitar uma análise comparativa de desempenho para as várias amostras de estudo pois, ambos os algoritmos permitem obter os mesmos valores de encaminhamento necessários ao cálculo de cada uma destas métricas. Em ambos se obtém uma estrutura de dados que permite, entre quaisquer dois pares de nós, obter não apenas as distâncias como também o número total de caminhos mínimos que passam ou não por um determinado nó.

Definida a estrutura de dados com a informação relevante de encaminhamento foram implementados em Java os algoritmos de cálculo das métricas de proximidade e intermediação.

Muito embora os dois algoritmos de encaminhamento escolhidos permitam obter os mesmos dados necessários ao cálculo das métricas de proximidade e de intermediação, eles são, contudo, algoritmos muito distintos na sua implementação. Desta forma, apresenta-se uma breve análise de cada uma destas adaptações implementadas que têm por base as implementações dos algoritmos originais.

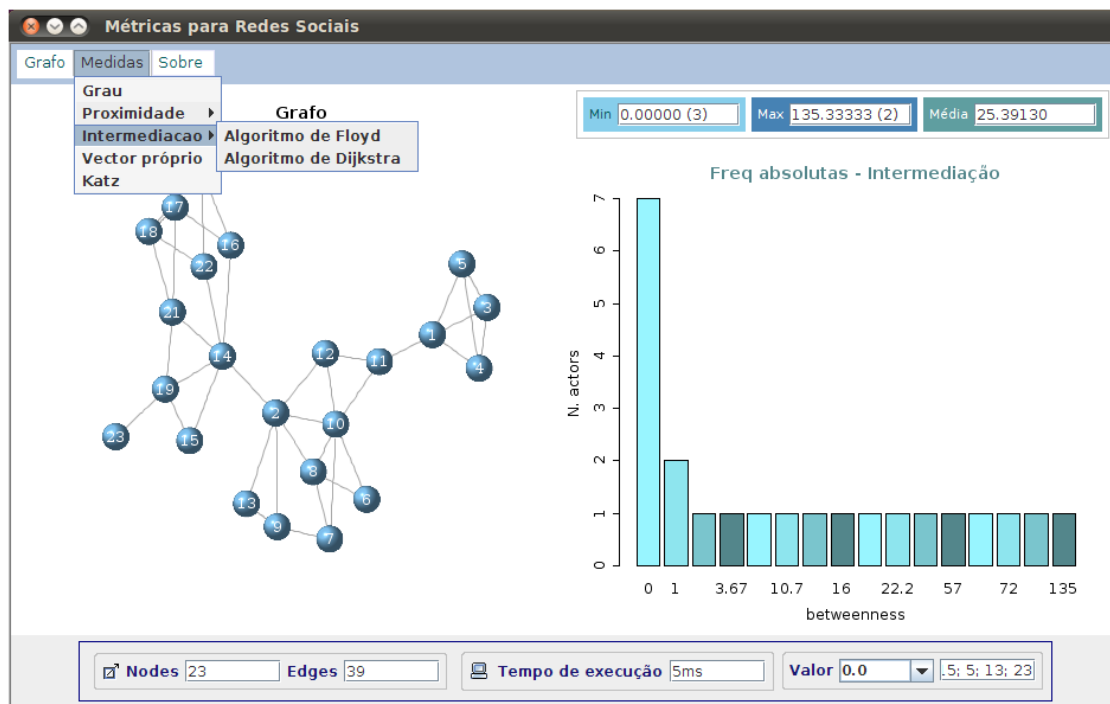


Figura 7 - Cálculo da Intermediação para a rede de estudo com 23 nós e 39 arestas

O algoritmo clássico de Floyd utiliza matrizes para determinar o peso dos caminhos mínimos entre todos os pares de nós. Neste são feitas n iterações que correspondem ao número total de nós da rede. A ideia básica deste algoritmo é verificar a cada iteração se a inclusão de um nó k intermédio no caminho de i para j pode reduzir o peso de um caminho já determinado. A cada iteração corresponde uma matriz $n \times n$ cujos valores são modificados utilizando a fórmula de recorrência:

$$d_{ij}^k = \min\{d_{ik}^{k-1} + d_{kj}^{k-1}, d_{ij}^{k-1}\} \tag{7}$$

onde d_{ij}^k é o peso mínimo entre os nós i e j na k -ésima matriz de iteração.

Estrutura do algoritmo de Floyd

1. Define-se a matriz D^0 , cujos valores d_{ij}^0 correspondem ao peso das arestas (i,j) se estas estão definidas ou $d_{ij} = \infty$ caso contrário. Convenciona-se que o peso de um nó definido para ele mesmo é 0 e assim a diagonal da matriz será $d_{ii} = 0$ para todo o i .
2. Para cada $k = 1 \dots n$ determinam-se sucessivamente os elementos da matriz D^k a partir dos elementos de D^{k-1} utilizando a expressão (7).
3. O processo do passo2 é repetido até $k = n$ e neste caso o valor do peso mínimo de todos os pares i,j do grafo estarão definidos na matriz D^n .

Na adaptação ao algoritmo proposta no trabalho desenvolvido, a matriz D^n será uma estrutura de dados que irá conter entidades que têm informação relativa não apenas às distâncias como também ao número total de caminhos mínimos, entre quaisquer dois pares de nós e ainda à lista de nós intermédios contidos nesses caminhos.

Tal como no algoritmo clássico, nesta adaptação são necessários três ciclos um para cada nova iteração de uma nova matriz e os outros dois para percorrer todas as posições da matriz. No entanto, como a matriz D^0 é simétrica, o 3º ciclo apenas percorre a triangular superior.

O algoritmo de Dijkstra assemelha-se ao algoritmo de busca em largura (Breadth-First Search - BFS), mas é um algoritmo guloso, ou seja, toma a decisão que parece ótima a cada momento. No algoritmo original, assim que é encontrado um caminho mínimo de A para B, os outros caminhos encontrados são descartados, mesmo que tenham o mesmo peso. Deste modo, esta foi uma entre muitas alterações que foi considerada na adaptação deste algoritmo, pois para calcular a medida da intermediação precisamos de saber o número total de caminhos mínimos entre A e B. De modo a garantir esta busca em largura otimizada, a ideia implementada consiste no seguinte: inicia-se a pesquisa no nó origem A e procuram-se todos os vizinhos deste nó, analisa-se cada um destes, e se algum for o nó destino, sabe-se que se encontrou um caminho. Todos os outros vizinhos podem ser desconsiderados, porque já não conseguem atingir o destino com um peso igual ou inferior. Na situação em que nenhuma das adjacências atinge o destino, constrói-se um novo caminho candidato para cada vizinho. Estes novos caminhos candidatos têm como pontos intermédios os nós do caminho anterior (que lhes deu origem).

Neste tipo de pesquisa é preciso garantir que não se entra em ciclo e para tal nesta implementação guardam-se as listas de intermédios que vão surgindo numa *hashtable* indexada por nome, e verifica-se em cada iteração se esse nome já é uma chave existente da *hashtable*.

Todos os candidatos a caminhos são guardados numa *LinkedList*, que tem referência direta para o primeiro e último nó da lista. Esta *LinkedList* desempenha o papel de um FIFO que guarda os candidatos. Sempre que se analisa um novo candidato, é escolhido o primeiro da lista. Sempre que se encontram novos candidatos eles são acrescentados ao final da lista. Desta forma, garante-se que a pesquisa é feita em largura. A implementação da *LinkedList* é a que mais se adequa a esta forma de uso, diminuindo a complexidade do algoritmo.

No início da análise de cada candidato, compara-se o seu peso com o peso mínimo de encaminhamento entre a origem e o destino naquela altura da execução. Se o presente candidato já tem um peso superior ao do caminho mínimo, ele pode ser imediatamente descartado.

6. Resultados Computacionais

O grafo considerado no caso de estudo é constituído por 23 nós e 39 arestas (vide Figura 8) e a instância que traduz esta rede tem o formato apresentado na Tabela 1. O ficheiro com extensão *.txt* ou *.csv* é composto por: comentários c (1ª linha), parâmetros p (2ª linha) número total de vértices (nodes) e número total de arestas (edges), conjunto de arestas (v_i, v_j) (a partir da 3ª linha) ordenado pelo vértice origem (v_i) .

Tabela 1 - Amostra k1-rede-23

c	nodes	edges
p	23	39
e	1	3
e	1	4
e	1	5
e	1	11
e	2	8
...

Segue-se a apresentação comparativa de resultados fornecidos pelo software apresentado e pelo NodeXL.

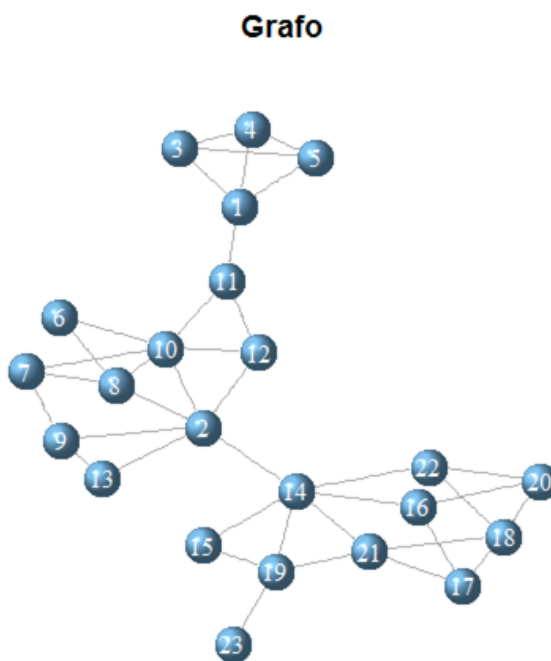


Figura 9 - Grafo k1-rede-23

6.1. Centralidade do Grau

Os resultados obtidos no *software* apresentado e no NodeXL foram unânimes. Os nós 2, 10 e 14 são os que apresentam o maior grau 6 (vide Figuras 9 e 10).

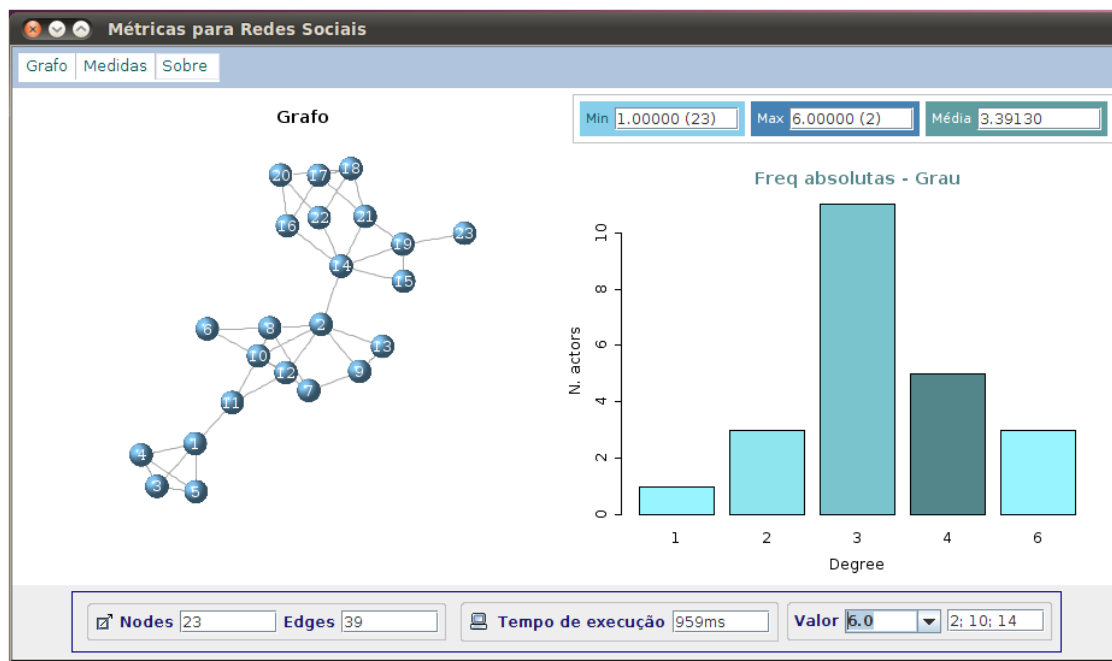
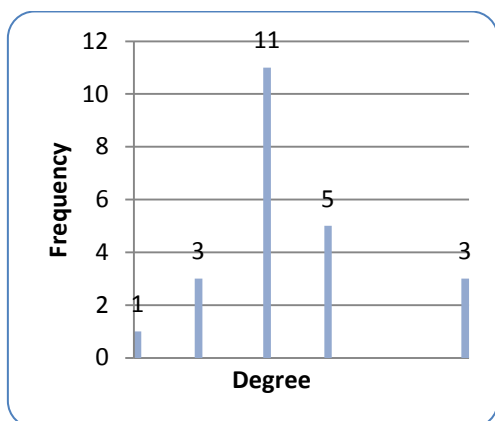


Figura 9 - Grau: Máximo grau 6 ao qual correspondem os nós 2,10 e 14

Todos estes nós, 2, 10 e 14, apresentam um maior número de ligações (6) com outros nós, podemos afirmar que são atores muito populares nesta rede social.



Minimum Degree	1
Maximum Degree	6
Average Degree	3,391
Median Degree	3,000

Figura 10 - Grau: Gráfico e rank obtido pelo NodeXL para a amostra k1-rede-23

No NodeXL o resultado gráfico não é tão evidente, mas é possível verificar que existem 3 nós que têm grau máximo 6. Sendo o NodeXL um aplicativo do programa Microsoft Excel é fácil comprovar na utilização do programa, efetuando uma consulta nas tabelas que os 3 nós são respetivamente o 2, o 10 e o 14.

6.2. Centralidade da Proximidade

A observação dos valores obtidos para o máximo em ambos os *softwares* permite concluir que apesar de um ligeiro erro de arredondamento, o nó 2 é o que apresenta maior valor de centralidade de proximidade (0.02040) (vide Figuras 11 e 12). Os resultados são mais uma vez concordantes.

A ligeira diferença nos gráficos deve-se ao facto do NodeXL efetuar um agrupamento de valores em classes com uma determinada amplitude.

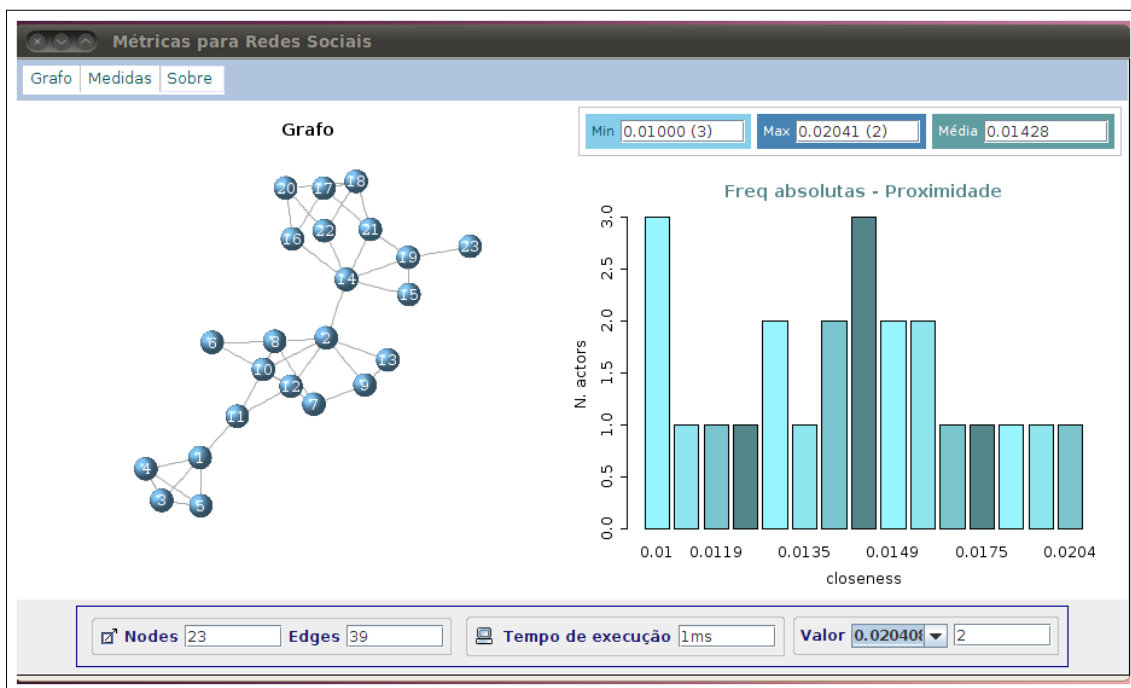
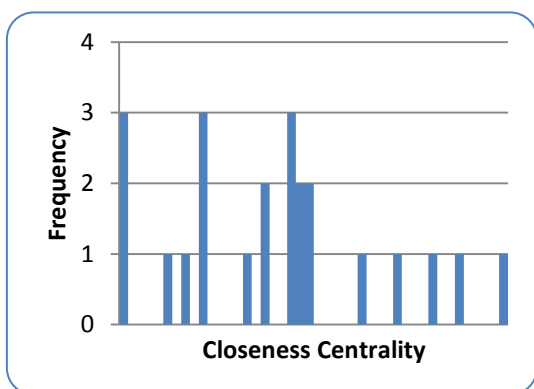


Figura 11 - Proximidade: Gráfico e *rank* para a amostra k1-rede-23



Minimum Closeness Centrality	0,010
Maximum Closeness Centrality	0,020
Average Closeness Centrality	0,014
Median Closeness Centrality	0,015

Figura 12 - Proximidade: Gráfico e *rank* obtido pelo NodeXL para a amostra k1-rede-23

Segundo esta medida o nó 2 é o ator mais próximo de todos os outros atores da rede, sendo aquele que apresenta a menor distância a todos os membros da rede.

6.3. Centralidade da Intermediação

Tal como na métrica anterior, a observação dos valores obtidos para o máximo em ambos os softwares permite concluir que o nó 2 é o que apresenta maior valor de centralidade de intermediação (135.(3)) (vide Figuras 13 e 14). Os resultados foram mais uma vez unânimes.

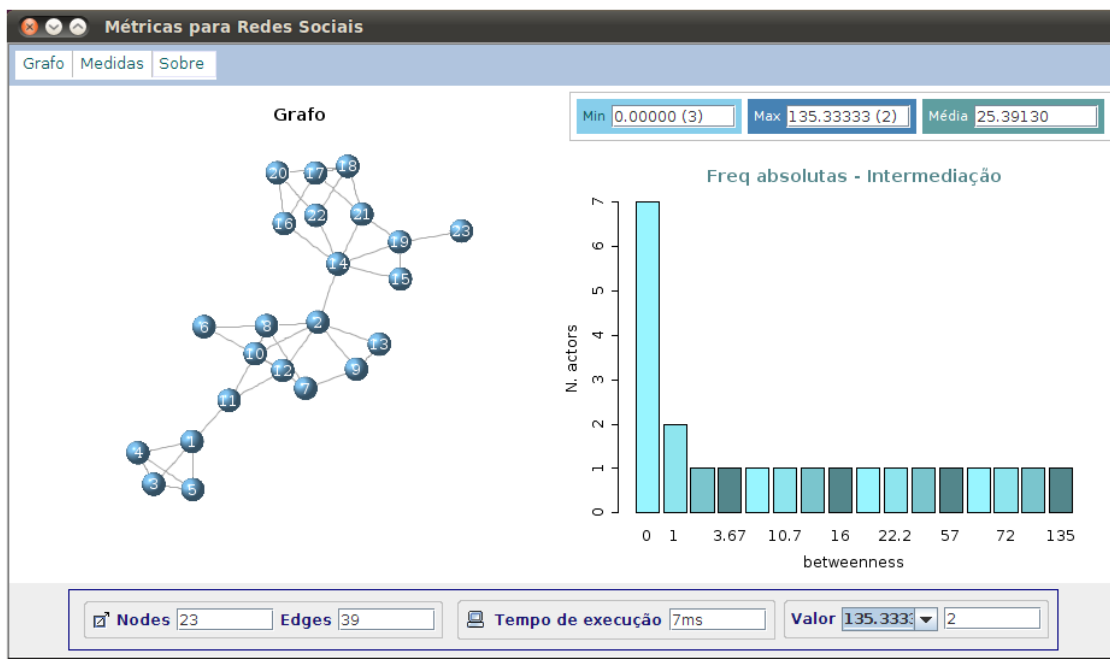


Figura 13 - Intermediação: Gráfico e *rank* para a amostra k1-rede-23

O nó 2 é um ator com intermediação máxima pois é um elo importante entre os vários grupos de nós e por ele circula grande parte do fluxo de informação entre os diferentes membros da rede. É caso para dizer que quanto mais atores dependem deste para estabelecer ligações mais poder este ator tem.

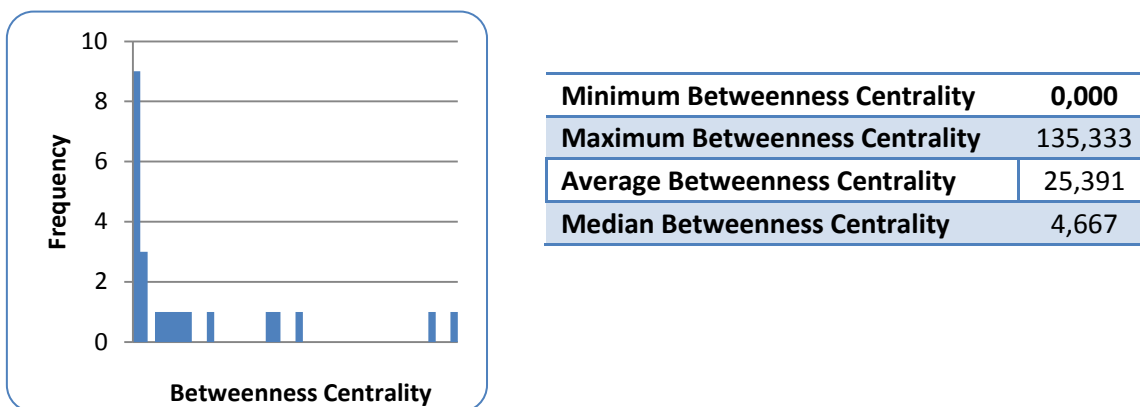


Figura 14 - Intermediação: Gráfico e *rank* obtido pelo NodeXL para a amostra k1-rede-23

6.4. Centralidade do vetor próprio

Ao contrário dos resultados anteriores, verifica-se uma diferença dos valores entre o *software* proposto e o NodeXL, visto que no NodeXL o valor do vetor próprio encontra-se normalizado. Enquanto o valor máximo para o NodeXL é de 0.106, o máximo para o *software* desenvolvido é 0.431 (vide Figuras 15 e 16). No entanto, ambos correspondem ao mesmo nó 2.

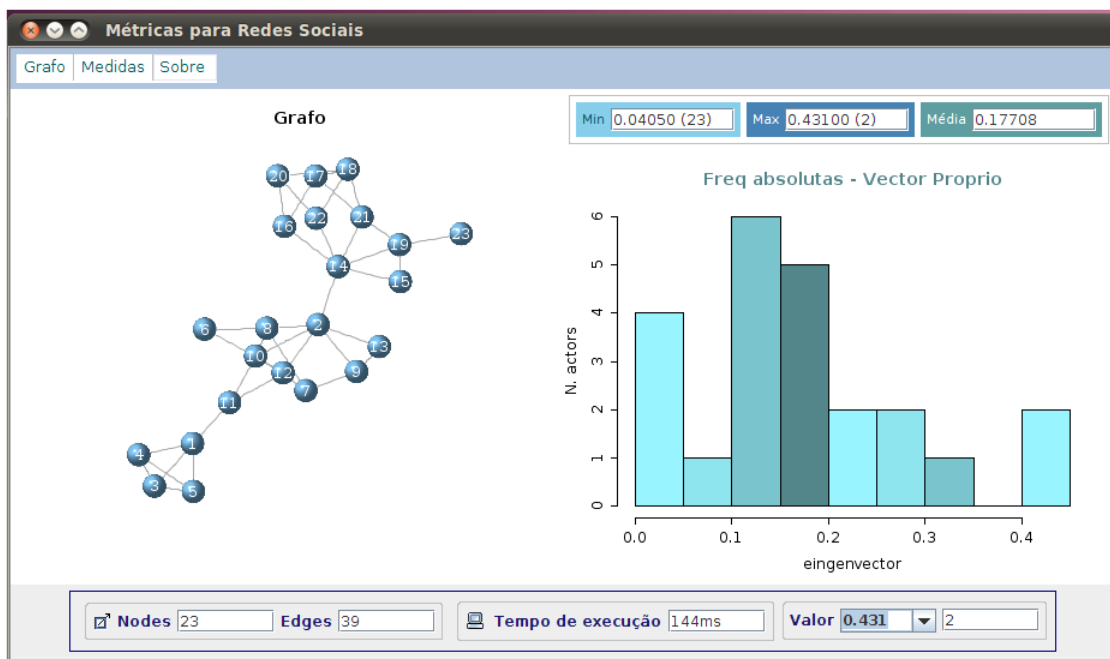
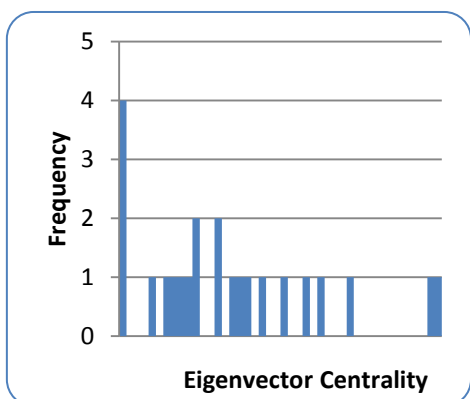


Figura 15 - Vetor próprio: Gráfico e rank para a amostra k1-rede-23



Minimum Eigenvector Centrality	0,010
Maximum Eigenvector Centrality	0,106
Average Eigenvector Centrality	0,043
Median Eigenvector Centrality	0,039

Figura 16 - Vetor próprio: Gráfico e rank obtido pelo NodeXL para a amostra k1-rede-23

A conclusão desta medida indica-nos que um nó importante tem vizinhos importantes, o ator 2 é importante porque depende de outros atores também eles importantes, é o caso, por exemplo dos vizinhos 8, 9,10, 12, 13 e do 14. Atores que têm mais ligações têm mais oportunidade porque têm mais escolhas.

6.5. Centralidade de Katz

O NodeXL não efetua o cálculo da métrica de centralidade de Katz, pelo que neste caso não é possível ter uma comparação. De qualquer modo, esta é apresentada a título de curiosidade e tal como na métrica anterior o nó 2 é o que tem maior valor de centralidade de katz (vide Figura 17).

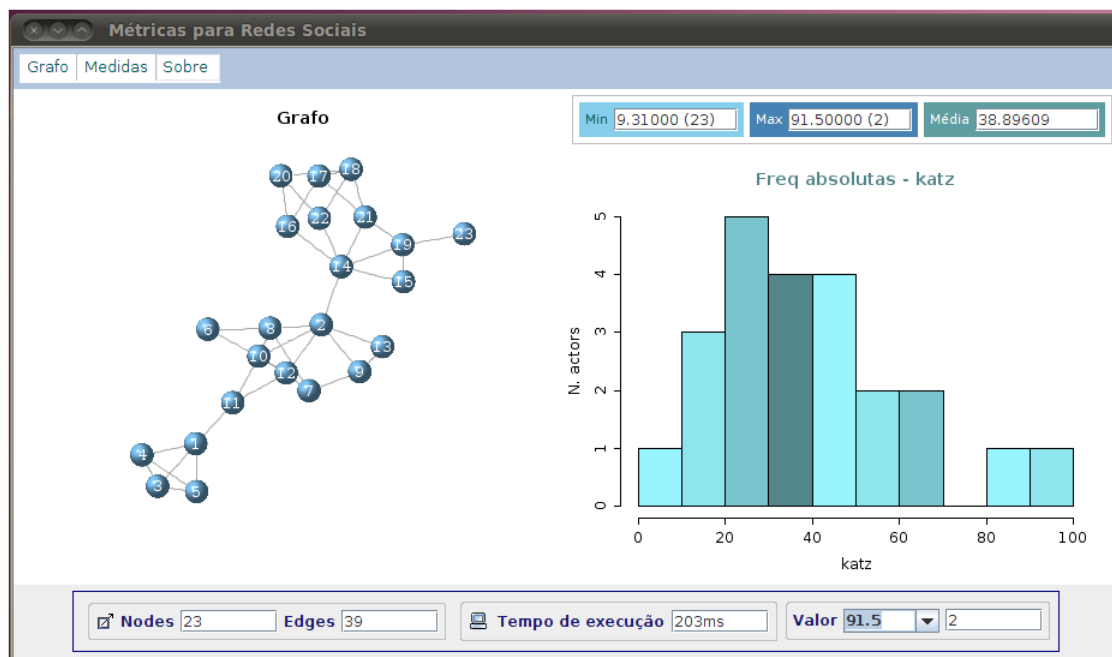


Figura 17 - Katz: Gráfico e rank para a amostra k1-rede-23

Nesta métrica, não apenas as ligações aos vizinhos diretos são importantes, mas também as interligações entre vizinhos dos vizinhos são importantes.

6.6. Síntese dos resultados computacionais

Na Tabela 2 apresenta-se a síntese comparativa dos resultados obtidos pelos dois softwares para a amostra em estudo e tecem-se algumas conclusões.

Tabela 2 - Síntese dos resultados da amostra k1-rede-23

Métricas	Software proposto	NodeXL
Grau	2, 10, 14	2, 10, 14
Proximidade	2	2
Intermediação	2	2
Vetor próprio	2	2
Katz	2	-

Ao analisar a Tabela 2 verifica-se que o ator número 2 tem certamente um papel fulcral nesta organização social. Este será sem dúvida um ator que dá coesão à rede e pelo qual o fluxo de informação circula chegando mais rápido e de forma mais abrangente a todos os outros atores da rede. Este poderá ser o ator importante na identificação de algum tipo de padrão que relacione os membros daquela organização.

Na métrica de grau são identificados outros atores importantes ou populares como é o caso do 10 e do 14. Estes atores são, em conjunto com o 2, os que apresentam mais ligações com os restantes elementos da rede, são portanto, os mais populares. Mas ao estudar outras métricas os atores 10 e 14 perdem a sua posição no ranking destacando-se apenas o ator 2.

O programa NodeXL não efetua o cálculo da métrica de Katz, desta forma não é possível tecer algum tipo de conclusão.

Noutras instâncias analisadas, as conclusões sobre um ou mais nós fulcrais na rede foram sempre concordantes para ambos os aplicativos (software desenvolvido e NodeXL).

Nalguns casos de estudo realizados demonstrou-se existirem mais do que um ator com importância ou poder na mesma rede. E diferentes medidas de centralidade apontavam diferentes atores como tendo um papel fundamental na rede. Isto significa que um ator importante para uma dada métrica de centralidade pode não ser tão crucial na análise de outras métricas de centralidade. Deste modo, as medidas em estudo devem ser analisadas em conjunto para poder formar-se algum tipo de conclusão pertinente à obtenção de informação que possa ser traduzida em conhecimento no estudo de uma dada rede social.

7. Conclusões

Neste artigo pretendeu-se dar a conhecer a importância e o modo de cálculo de algumas métricas de centralidade aplicadas a redes sociais. A implementação do *software* permite calcular cinco métricas de centralidade.

O *software* foi construído recorrendo a duas linguagens de programação: o Java e o *software* de computação estatística R. Com o uso do programa R, o cálculo de algumas métricas acabou por se tornar muito vantajoso, ao invocar os scripts de R através do Java. As métricas de proximidade e de intermediação foram implementadas em Java porque usavam dois algoritmos de cálculo de caminhos mínimos distintos.

No que diz respeito à comparação de resultados obtidos entre as métricas determinadas pelo software dado e o NodeXL, as conclusões não deixam dúvidas, os resultados sobre os nós são unânimes em ambos os programas. No caso da métrica de centralidade de vetor próprio, para a qual foram obtidos valores diferentes, se alterarmos o algoritmo que implementa o cálculo desta métrica de modo a obter a normalização dos vetores próprios teremos mais uma vez, tal como aconteceu em todas as outras métricas, valores iguais em ambos os softwares.

A versão desenvolvida é apenas um protótipo 1.0 que ainda não foi batizado e que poderá ser bastante aperfeiçoado num trabalho futuro. Uma das grandes alterações deverá passar por poder usar o programa em todos os sistemas operativos, pois neste momento apenas é possível usar em Linux; ou ainda, disponibilizá-lo na Web. Outras alterações poderão passar por tornar o ambiente gráfico mais interativo e testá-lo com grandes volumes de dados. A inclusão de outras métricas, como a PageRank, também está nos nossos horizontes.

Referências

- Barabási, A.-L., Albert, R. (1999) Emergence of scaling in random networks, *Science*, vol. 286 (5439), pp. 509–512.
- Bonacich P., Lloyd P., (2001) Eigenvector-like measures of centrality for asymmetric relations, *Social Networks*, vol. 23(3), pp. 191-201.
- Dijkstra, E.W.(1959) A note on two problems in connection with graphs, *Numerische Mathematik*, vol. 1, pp. 269–271.
- Ellison, N., Steinfield, C., Lampe, C. (2007) The benefits of Facebook “friends:” social capital and college students' use of online social network sites, *Journal of Computer-Mediated Communication*, vol. 12, pp. 1143–1168.
- Erdos, P., Renyi, A. (1959) On Random Graphs. I., *Publicationes Mathematicae*, vol. 6, pp. 290–297.
- Floyd, R.W. (1962) Algorithm 97: Shortest Path, *Communications of the ACM*, vol. 5 (6) pp. 345.
- Gama, J., A.P.L. Carvalho, M. Oliveira, A.C. Lorena, K. Faceli (2012) *Extração de Conhecimento de Dados/Data Mining*, Lisboa, Edições Sílabo.
- Johnson, S. (2001) *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. Allen Lane.
- Landherr, A., Friedl, B., Heidemann, J. (2010) A Critical Review of Centrality Measures in Social Networks, *Business & Information Systems Engineering* vol. 2(6), pp. 371–85.
- Katz, L. (1953) A New Status Index Derived from Sociometric Index, *Psychometrika* vol. 18(1), pp. 39-43.
- Milgram, S., (1967) The small world problem, *Psychology Today*, vol. 2, pp. 60-67.
- Travers, J. and Milgram, S. (1969) An experimental study of the small world problem, *Sociometry*, vol. 32, pp. 425-443.
- Moreno, J. L. (1934) *Who Shall Survive?*, Beacon House, Beacon, NY.
- Rapoport, A. (1957) Contribution to the theory of random and biased nets, *Bulletin of Mathematical Biophysics*, vol. 19, pp. 257-277.
- Scott, J.(2000) *Social Network Analysis - A Handbook*, Sage Publications, London.
- Sabidussi, G. (1966) The centrality index of a graph, *Psychometrika*, vol. 31(4), pp. 581-603.

Watts, D.J., Strogatz, S.H. (1998) Collective dynamics of small-world networks, Nature, vol. 393(6684), pp. 409–410.

Watts, D. J. (2003) Six Degrees: The Science of a Connected Age. Norton.



Paula Alexandra Laranjeira é licenciada em Informática pela Universidade Aberta e licenciada em Ensino de Matemática pela Universidade de Aveiro. É também pós-graduada em Inovação e Criatividade: Tecnologias da Informação e Comunicação pelo ISCE. Foi produtora de conteúdos multimédia na empresa Porto Editora, estando diretamente envolvida no projeto Escola Virtual. Foi *web designer* na Sociedade Ibérica de Citometria. Atualmente está ligada à coordenação e formação de projetos e estágios de cursos profissionais de informática e à certificação de jovens e adultos.



Luís Cavique, Professor Auxiliar no Departamento de Ciências e Tecnologia (DCeT), Secção de Informática, Física e Tecnologia (SIFT). Coordenador da Licenciatura em Informática no biénio 2012 - 2014. Licenciado em Engenharia Informática em 1988 pela FCT-UNL. Obteve o grau Mestre em Investigação Operacional e Eng. Sistemas pelo IST-UTL em 1994. Obteve o grau de Doutor em Eng. Sistemas pelo IST-UTL em 2002. Tem como áreas de interesse, a intersecção da Informática (Computer Science) com a Engenharia de Sistemas (Management Science) designadamente a área de “Data and Graph Mining”.