

Técnicas de análise e visualização com grafos na Psicologia: utilização do NodeXL

Graph analysis and visualization techniques in Psychology: using NodeXL

Técnicas de análisis y visualización con gráficos en Psicología: uso de NodeXL

Recebido: 17/06/2020 | Revisado: 01/07/2020 | Aceito: 06/07/2020 | Publicado: 20/07/2020

Maély Ferreira Holanda Ramos

ORCID: <https://orcid.org/0000-0001-6150-6345>

Universidade Federal do Pará, Brasil

E-mail: maelyramos@hotmail.com

Fernando Augusto Ramos Pontes

ORCID: <https://orcid.org/0000-0001-9569-943X>

Universidade Federal do Pará, Brasil

E-mail: fernando.pontes@pesquisador.cnpq.br

Simone Souza da Costa Silva

ORCID: <https://orcid.org/0000-0003-0795-2998>

Universidade Federal do Pará, Brasil

E-mail: symon.ufpa@gmail.com

Erika Cristina de Carvalho Silva Pereira

ORCID: <https://orcid.org/0000-0003-0701-9815>

Universidade Federal do Pará, Brasil

E-mail: erika7carvalho@gmail.com

Resumo

A revisão sistemática consiste em método essencial para o levantamento de dados sobre determinado objeto de estudo. É um tipo de pesquisa que requer um protocolo rígido de busca, coleta, análise e apresentação dos dados, podendo ou não utilizar métodos estatísticos. Nesse sentido, a Teoria dos Grafos é um dos ramos da matemática que oferece importantes recursos para a análise e representação das relações entre variáveis de qualquer natureza, sendo muito útil também em pesquisas como as revisões sistemáticas na área da psicologia. O presente artigo constitui-se em uma proposta metodológica inovadora que visa descrever: o processo básico de construção, interpretação e a utilização de grafos em revisões sistemáticas aplicadas à área da psicologia. Os resultados indicam que com utilização das técnicas de pesquisa com grafos é possível visualizar e identificar regras de associação entre palavras-

chave, por exemplo, além de auxiliar na investigação das relações entre as variáveis em estudo. Conclui-se que as técnicas de análise e visualização por meio de grafos, bem como a aplicação de medida de centralidade, demonstraram-se adequadas e eficazes, podendo contribuir com estudos de revisão sistemática.

Palavras-Chave: Teoria dos grafos; Revisão sistemática; Psicologia

Abstract

The systematic review is an essential method for collecting data on a particular object of study. It is a type of research that requires a strict protocol for searching, collecting, analyzing and presenting data, with or without the use of statistical methods. In this sense, Graph Theory is one of the branches of mathematics that offers important resources for the analysis and representation of the relationships between variables of any nature, being also very useful in research such as systematic reviews in the field of psychology. This article is an innovative methodological proposal that aims to describe: the basic process of construction, interpretation and the use of graphs in systematic reviews applied to the area of psychology. The results indicate that with the use of search techniques with graphs it is possible to view and identify rules of association between keywords, for example, in addition to assisting in the investigation of the relationships between the variables under study. It is concluded that the techniques of analysis and visualization by means of graphs, as well as the application of a centrality measure, proved to be adequate and effective, being able to contribute to systematic review studies.

Keywords: Graph theory; Systematic review; Psychology

Resumen

La revisión sistemática es un método esencial para recopilar datos sobre un objeto particular de estudio. Es un tipo de investigación que requiere un protocolo estricto para buscar, recopilar, analizar y presentar datos, con o sin el uso de métodos estadísticos. En este sentido, Teoría de Grafos es una de las ramas de las matemáticas que ofrece recursos importantes para el análisis y la representación de las relaciones entre variables de cualquier naturaleza, siendo también muy útil en investigaciones como las revisiones sistemáticas en el campo de la psicología. Este artículo es una propuesta metodológica innovadora que tiene como objetivo describir: el proceso básico de construcción, interpretación y uso de gráficos en revisiones sistemáticas aplicadas al área de la psicología. Los resultados indican que con el uso de técnicas de búsqueda con gráficos es posible ver e identificar reglas de asociación entre

palabras clave, por ejemplo, además de ayudar en la investigación de las relaciones entre las variables en estudio. Se concluye que las técnicas de análisis y visualización mediante gráficos, así como la aplicación de una medida de centralidad, demostraron ser adecuadas y efectivas, pudiendo contribuir a estudios de revisión sistemática.

Palabras clave: Teoría de grafos; Revisión sistemática; Psicología

1. Introdução

A revisão sistemática da literatura é um método fundamental para quem busca evidências sobre determinado assunto, sendo um método moderno para a avaliação simultânea de conjuntos de dados, pois reuni os conhecimentos fracionados em diferentes estudos sobre os constructos de interesse de qualquer área, possibilitando a construção de panoramas com rigor acadêmico e confiabilidade de informações (Castro, 1992; Cordeiro et al., 2007).

Uma revisão eficiente deve ser planejada sistematicamente, respondendo à uma pergunta específica, utilizando métodos explícitos para identificar, selecionar e avaliar criticamente os estudos envolvidos na pesquisa. As revisões sistemáticas são consideradas estudos originais, pois utilizam como fontes de informação diferentes dados disponibilizados na literatura (Castro, 1992; Cordeiro et al., 2007; Rother, 2007; Sousa & Ribeiro, 2009).

Como um tipo de investigação científica, a revisão sistemática possibilita reunir, avaliar criticamente e sintetizar resultados de múltiplos estudos para o aprofundamento da compreensão de diferentes temáticas, podendo ou não englobar métodos estatísticos (metanálise). Exige-se um protocolo rígido para a busca, coleta, análise e apresentação de dados (Castro, 1992; Cordeiro et al., 2007; Rother, 2007; Sousa & Ribeiro, 2009).

Todo este cuidado com o desenvolvimento da revisão sistemática se justifica na necessidade de se evitar tendenciosidades (viés), minimizando erros nas conclusões. Assim, considera-se que este tipo de estudo permite a apropriação das melhores evidências sobre determinado assunto. Ao final de uma revisão sistemática é possível, portanto, obter uma visão geral sobre o fenômeno investigado (Castro, 1992; Cordeiro et al., 2007; Rother, 2007; Sousa & Ribeiro, 2009). Desta forma, entende-se que a utilização de ferramentas capazes de auxiliar os pesquisadores na análise dos dados e na apresentação dos resultados das revisões sistemáticas, pode aprimorar este método de estudo, tornando-o ainda mais eficiente. Uma das possibilidades é a utilização de técnicas de análise e visualização com grafos, que visem

representar por meio de diagramas os resultados dos estudos, podendo ser importantes alternativa para este tipo de pesquisa.

Considera-se que esta Teoria dos Grafos foi criada pelo matemático Euler em 1736, tendo sua origem na problematização de uma situação recreativa, conhecida como as *Sete Pontes de Königsberg*. O problema consistia em verificar se era possível dar um passeio atravessando cada uma das pontes que passavam sobre o rio Pregel, uma única vez, e voltando ao ponto de partida, na cidade de Königsberg (Prússia). Assim, Euler criou um esquema representando as quatro partes da cidade através de quatro pontos, unindo estes pontos por linhas que representavam as pontes. Ao descrever desta forma o problema das pontes de Königsberg, Euler deu início ao que hoje se conhece por Teoria dos Grafos (Matos, 2013; Pinheiro, 2013; Sousa, 2010).

A Teoria dos Grafos indica que diferentes relações entre objetos de qualquer natureza podem ser representadas por meio de grafos. Entende-se que pode ser aplicada em qualquer área do conhecimento, como, neste caso, na área da psicologia. Os grafos são fundamentados em técnicas e recursos computacionais e matemáticos para caracterizar a estrutura topológica de relações (Matos, 2013; Pinheiro, 2013; Sousa, 2010). Considera-se que a utilização de grafos para investigar e representar as conexões entre as variáveis investigadas nas revisões sistemáticas podem representar importante auxílio à pesquisadores.

Diferentes problemas de interesses acadêmicos e práticos podem ser investigados utilizando grafos. Esse tipo de estudo e representação pode ser feito por programas específicos de técnicas de análise de rede com grafos, tais como o NodeXL. O NodeXL, versão 1.0.1.251, é um modelo gratuito, servindo como um complemento *open-source* para o Excel 2007, 2010 e 2013, que dispõe de uma interface gráfica para a interação com o usuário. Através do NodeXL é possível calcular métricas para identificar as interações dinâmicas, regras de intermediação e graus de centralidade entre os elementos que compõem o conjunto de variáveis de interesse do pesquisador, bem como gerar representações gráficas, em forma de diagramas, resultantes das relações investigadas (Hansen et al., 2010; Smith, 2013).

Desta forma, o presente artigo tem como meta contribuir metodologicamente com revisões sistemáticas da área da psicologia, descrevendo a utilização da Teoria dos Grafos, a partir da ferramenta NodeXL, para aplicação em estudos dessa natureza. Tem como objetivos: (1) descrever o processo básico de construção e interpretação de grafos; (2) descrever a utilização de grafos com o NodeXL, em revisões sistemáticas.

2. Metodologia

Este artigo é um estudo metodológico que apresenta a utilização de grafos em pesquisas de revisão sistemática, como recurso exploratório de análise e de visualização dos dados. Para tanto foi realizada a demonstração da ferramenta NodeXL na construção de grafos e suas possibilidades de aplicação em estudos de revisão sistemática. Desta forma cumpriram-se as seguintes etapas: (1) apresentação dos conceitos básicos de grafos; (2) interpretação de métricas; (3) aplicação de grafos com o NodeXL em revisão sistemática.

Nas etapas 1 e 2 foram apresentados conceitos e orientações básicas para interpretar de grafos e métricas. Consideraram-se as seguintes medidas: (1) *Coefficiente de Agrupamento/C(G)* – para gerar grupos; (2) *Grau de Centralidade (in-degree/out-degree)* – para quantificar o número de ligações associadas à um elemento; (3) *Centralidade de intermediação (Betweenness Centrality)* - para verificar o grau de participação de um vértice nos caminhos mais curtos de um grafo; (4) *Centralidade de intermediação (Eigenvector Centrality)* - para medir o grau de importância de um vértice (5) *Centralidade de Proximidade (Closeness Centrality)* – para calcular quão próximo está cada vértice dos demais; (6) *Relevância do Nó (PageRank)* – para medir o nível de importância de um vértice em grafos dirigidos; (7) *Densidade (Density)* – para verificar a coesão global (Matos, 2013; Pinheiro, 2013; Sousa, 2010). Na terceira etapa foram demonstrados os procedimentos necessários para a construção de grafos na ferramenta NodeXL.

Na terceira etapa apresentou-se algumas possibilidades de análise de dados e representação dos resultados, mediante identificação de regras de associação entre palavras-chaves de artigos e no reconhecimento das relações entre os resultados de estudos, desenvolvendo representações geométricas. Para demonstração dessas técnicas analisou-se dados (palavras-chave e principais variáveis investigadas pelos estudos) que correspondem à 30 artigos selecionados a partir do diretório da Capes-Periódicos, no período entre 2000 a 2013.

Os artigos foram selecionados a partir da seguinte estratégia de buscas: (*job satisfaction OR teacher satisfaction OR work satisfaction*) AND (*school OR educational institution*) AND *teacher* AND (*teacher self-efficacy OR self-efficacy OR teacher efficacy OR beliefs efficacy*) AND (*teacher collective efficacy OR collective efficacy*) AND (*theory social cognitive OR social cognitive theory*).

Visando o delineamento da pesquisa, foram pré-determinados alguns critérios de inclusão, sendo esses: conter no título, no resumo, ou/e nas palavras-chave os termos *job*

satisfaction e teacher; estar completo e disponível; apenas artigos dos últimos 13 anos (2000 a 2013); artigos em inglês ou português; ter sido revisado por pares; e ter como foco investigar o constructo da satisfação no trabalho docente, utilizando os conceitos da Teoria Social Cognitiva de Bandura (1986). Todos os achados que não obedeceram aos critérios de inclusão mencionados foram excluídos.

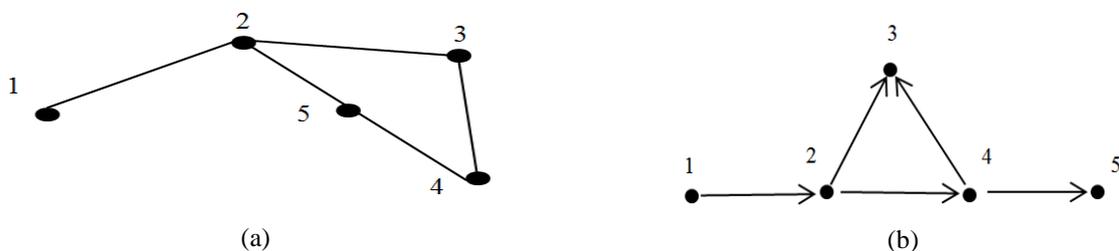
3. Resultados e Discussão

Conceitos básicos de grafos

Um grafo $G = (V, E)$ consiste num conjunto finito e não vazio V de n vértices e um conjunto E de m arestas, formado por pares de elementos distintos de V . Considera-se os elementos V como um conjunto de vértices e os elementos E como um conjunto de arestas. Os vértices V representam elementos de qualquer fenômeno, que são ligados por arestas E que indicam qualquer relação entre os vértices V . Assim, pode-se entender que os vértices V são os elementos e as arestas E são as conexões entre eles. A aresta $e \in E$ é denotada pelo par de vértices $e = (v_1, v_2)$ que a forma. Considera-se que os vértices $(v_1, v_2) \in V$ são os extremos da aresta (e), e por isso são denominados adjacentes ou vizinhos (Szwarcfiter, 1984; Freitas, 2010).

Pode-se visualizar um grafo a partir de sua representação geométrica, na qual seus vértices correspondem a pontos e as arestas são associadas à linhas arbitrárias que unem os pontos, nesse sentido, o conjunto de arestas E induz uma relação binária em V (Szwarcfiter, 1984; Freitas, 2010).

Figura 1 – Exemplo de grafo não direcionado e cíclico (a) e de grafo direcionado (b).



Fonte: Elaboração dos autores (2020).

As Figuras 1(a) e 1(b) são representações geométricas de grafos $G = (V, E)$, onde os vértices da Figura 1(a) são: $V = \{1, 2, 3, 4, 5\}$; e as arestas são $E = \{(1,2), (2,3), (2,5), (3,4), (4,5)\}$. Os vértices da Figura 1(b) são: $V = \{1, 2, 3, 4, 5\}$; e as arestas são $E = \{(1,2), (2,3), (2,4), (4,3), (4,5)\}$. Ao observar inicialmente um grafo é necessário identificar se são direcionados (no caso representam relações não simétricas entre os elementos) ou não direcionados (para relações simétricas). A Figura 1(a) é um grafo não direcionado, desta forma, a aresta (v_1, v_2) é incidente a ambos v_1 e v_2 . Define-se que o grau é o número de arestas que incidem sobre o vértice. Por exemplo, o grau do vértice 2 no grafo da Figura 1(a) é igual a 3. Nos grafos direcionados, também denominados de dígrafos, cada aresta (v_1, v_2) possui uma direção indicada por uma seta (\longrightarrow), podendo ser divergente de v_1 (quando a seta parte de v_1) e convergente de v_2 (quando v_2 recebe a seta), ou o contrário. A Figura 1(b) é um exemplo de grafo direcionado. Deve-se considerar que o grau de entrada de v_1 é o número de arestas convergentes à v_1 . Já o grau de saída de v_1 é o número de arestas divergentes de v_1 . Por exemplo, o grau de entrada e saída do vértice 4 no grafo da Figura 1(b) é igual a 1 e 2, respectivamente (Szwarcfiter, 1984).

O caminho num grafo é uma sequência de vértices $v_1 \dots v_k$. Se todos eles forem distintos constitui-se um caminho simples ou elementar. No entanto, se as arestas forem distintas então denomina-se a sequência de trajeto. Frisando que, em dígrafos, o caminho deve obedecer ao direcionamento das arestas. O comprimento de um caminho equivale à quantidade de arestas presentes no caminho. Os grafos também podem apresentar ciclos. Considera-se *cíclico* um grafo que possui ao menos um caminho fechado entre seus vértices. Isto significa que o vértice inicial será também o final (v_1, v_2, \dots, v_1) , seguindo uma sequência (Szwarcfiter, 1984). Por exemplo, o caminho $\{2, 3, 4, 5, 2\}$ de comprimento 4 no grafo da Figura 1(a) configura-se um ciclo.

Busca-se frequentemente nos grafos perceber visualmente a aproximação entre os elementos envolvidos no estudo. A *vizinhança* entre dois vértices se dá quando ambos os elementos se conectam à uma mesma aresta. Considera-se que dois vértices v_1 e $v_2 \in V$ são adjacentes ou vizinhos quando existe uma aresta $e \in E$, sendo que $e = (v_1, v_2)$. Desta forma, a vizinhança de um vértice $v \in V$ consiste em todos os vértices pertencentes à V que se conectam à v por uma aresta (Szwarcfiter, 1984).

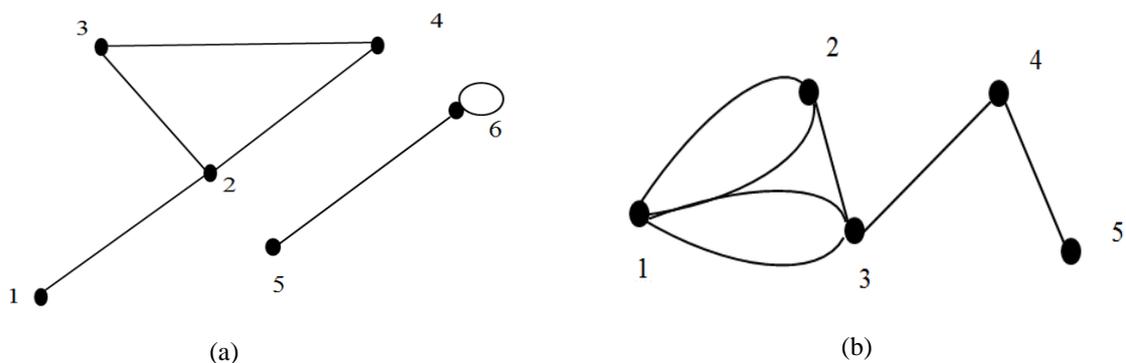
É preciso avaliar se a representação geométrica apresenta um grafo simples ou um multigráfo. Grafos simples não possuem arestas paralelas, nem laços. Quando o grafo possui arestas paralelas, ou seja, mais de uma aresta entre o mesmo par de vértices, o mesmo é

denominado multigráfo (Figura 2 (b)). O laço é representado por uma aresta que liga um vértice a si mesmo, do tipo (v_i, v_i) , por exemplo, a aresta (6, 6) presente no grafo da Figura 2(a) (Szwarcfiter, 1984).

Os grafos podem ser caracterizados também como conexos e desconexos. Quando um grafo $G = (V, E)$ é conexo significa que existe pelo menos um caminho entre cada par de vértice. Quando não é possível identificar pelo menos um caminho entre qualquer par de vértices, ele é considerado desconexo, sendo sua representação descontínua. No entanto, quando um grafo não possui arestas, isto significa que é totalmente desconexo (Szwarcfiter, 1984).

Considera-se que grafos podem possuir subgrafos, entendendo que um *subgrafo* $G' = (V', E')$ de um grafo $G = (V, E)$ é um grafo em que $V' \subset V$ e $E' \subset E$. Ressalta-se que subgrafos podem formar cliques, que são subconjuntos completos de um grafo, onde cada elemento está diretamente ligado à todos os demais, demonstrando total coesão global, a partir da existência de uma aresta para cada par de vértices distintos. O tamanho de um clique é indicado pela cardinalidade de seu conjunto de vértices (Szwarcfiter, 1984).

Figura 2 - Exemplo de grafo desconexo, com clique e laço (a) e um multigráfo conexo (b).



Fonte: Elaboração dos autores (2020).

A Figura 2(a) representa um grafo G com um clique de tamanho 3, composto pelos vértices $\{2, 3, 4\}$. Neste exemplo, o subgrafo G' , com $V' = \{2, 3, 4\}$, é formado por três vértices interligados, constituindo um subconjunto completo de G .

Interpretação de métricas

Para interpretar corretamente os grafos é necessário compreender a forma como são criados. As métricas são importantes para classificar os vértices e posicioná-los nos grafos,

sendo medidas quantitativas que caracterizam vários aspectos de um grafo, indicando graus de centralidade, intermediação, proximidade e relevância, entre outros (Hansen et al., 2011; Hansen et al., 2010; Smith, 2013).

Principais métricas

Sustentando-se nos conceitos provenientes da Teoria dos Grafos apresentam-se medidas de centralidade que são importantes para representar as relações entre os vértices, a saber: (1) *degree*; (2) *Betweenness Centrality*; (3) *Closeness Centrality*; (4) *Eigenvector Centrality*; (5) *PageRank*; (6) *Clustering Coefficient*; (7) *Density*.

A centralidade de grau (*degree*), é a contagem do número de arestas que incidem sobre determinado vértice, quanto mais arestas um par de vértices receber, mais forte será a conexão entre eles. Em grafos direcionados o grau de centralidade pode ser dividido em dois tipos: *in-degree*, que quantifica o número de arestas que convergem para um vértice de interesse (número de arestas incidentes do nó); *out-degree*, que mede o número de arestas emergentes do vértice de interesse (Matos, 2013; Pinheiro, 2013; Sousa, 2010; Freitas, 2010).

A centralidade de intermediação (*Betweenness Centrality*) é medida levando-se em consideração a ligação entre vértices vizinhos, onde vértices que se configuram como ligação entre indivíduos/grupos recebem maiores escores de intermediação. Geralmente vértices que são incluídos em mais caminhos curtos entre outros vértices, tem maior capacidade de intermediação. Assim, esta métrica define o grau de participação de um vértice nos caminhos mais curtos de um grafo (Matos, 2013; Pinheiro, 2013; Sousa, 2010).

A centralidade de proximidade (*Closeness Centrality*) indica a posição mais central dos vértices, medindo a proximidade entre um elemento e os outros do grafo. Baseia-se na soma das distâncias de um vértice em relação a todos os demais, quando um vértice é central ele possui a capacidade de interagir rapidamente com os outros. Considera-se que a distância entre dois vértices de um grafo é o comprimento do menor caminho entre eles. Assim, os escores são normalizados para se obter valores entre 0 e 1. Quanto maior for o escore, maior será a capacidade de aproximação (Szwarcfiter, 1984; Matos, 2013; Pinheiro, 2013; Sousa, 2010; Freitas, 2010).

A centralidade de autovetor (*eigenvector centrality*) leva em consideração não só o número de conexões que um vértice tem, mas também os graus de centralidade dos outros vértices com os quais está conectado. Assim, através dessa métrica mede-se a importância de um elemento na rede de relações, atribuindo valores a todos os vértices. Um valor elevado

nesta métrica significa que um elemento tem uma posição central na rede e que também está conectado a outros elementos centrais (Matos, 2013; Pinheiro, 2013; Sousa, 2010).

A métrica *PageRank*, é aplicada, apenas, a grafos direcionados. Suas medidas indicam que um vértice é mais conectado aos demais do grafo e, que por isso, é considerado importante. Espera-se que conteúdos de vértices importantes sejam mais significativos que os conteúdos de vértices adjacentes (periféricos). Por sua capacidade de encontrar os vértices mais importantes o *PageRank* assemelha-se à métrica *eigenvector centrality*. Desta forma, pode-se definir o *PageRank* como o grau de importância do vértice v , calculada pela soma da importância dos vértices que apontam para v (Matos, 2013; Pinheiro, 2013; Sousa, 2010).

O coeficiente de agrupamento $C(G)$ (*Clustering Coefficient*) é uma métrica gerada pela medida de valor relativo de triângulos existentes no grafo. Seu valor é calculado a partir da média dos coeficientes de agrupamento de todos os vértices. Para tanto, calcula-se o coeficiente de agrupamento de cada vértice do grafo, estimando posteriormente a média das somas desses valores. Considera-se que quanto maior for o $C(G)$ de um grafo, mais curtos serão os caminhos entre os seus vértices (Matos, 2013; Pinheiro, 2013; Sousa, 2010).

O grau de densidade (*density*) de um grafo indica a coesão global de uma rede de relações, sendo calculada pela soma de ligações realizadas pelos elementos que compõem o grafo, este número é posteriormente dividido pela quantidade de possível relações que poderiam ser estabelecidas. Assim a coesão global de um grafo é quantificada em escores entre os intervalos 0 e 1. Quanto mais próximo de 1 o grafo estiver, mais coeso e completo ele será. A densidade de um grafo está relacionada à quantidade de ligações realizadas, quanto mais abundantes forem, mais densa será a rede (Matos, 2013; Pinheiro, 2013; Sousa, 2010).

Aplicação de grafos em revisão sistemática para a área da psicologia

Para realizar a demonstração de técnicas de análise e visualização com grafos através do NodeXL, utilizou-se dados resultantes da seleção de 30 artigos, em inglês, relacionados à satisfação no trabalho docente, a partir da Teoria Social Cognitiva, no diretório da CAPES-Periódicos.

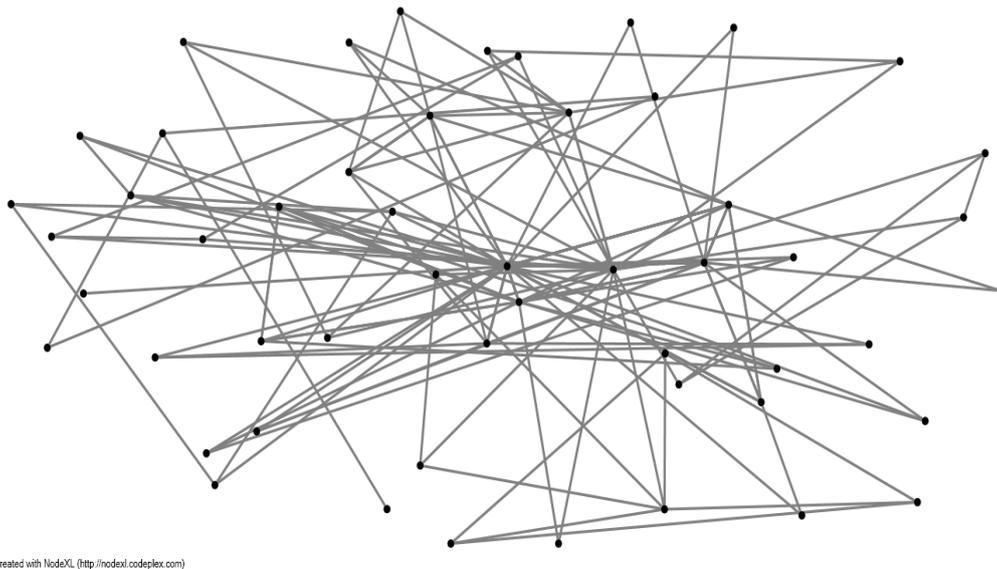
A satisfação no trabalho é uma condição prazerosa de um estado emocional positivo, resultante da avaliação de experiências de trabalho. Pode também ser considerada como julgamentos positivos resultantes das avaliações que as pessoas fazem acerca dos seus trabalhos (Duffy & Lent, 2009; Locke, 1976).

Enfatiza-se que o objetivo deste tópico não é explorar os constructos relacionados à temática, mas sim ilustrar as técnicas de análise com grafos, sinalizando que outros pesquisadores podem utilizar técnicas semelhantes na exploração dos seus dados em estudos de revisão sistemática.

Descrição de técnicas de análise com grafos para a identificação de regras de associação entre palavras-chave

São apresentadas neste tópico formas de utilização de grafos gerados no NodeXL para investigação de possíveis regras de associação entre as palavras-chave de artigos selecionados para estudos de revisão sistemática. A Figura 3, representa a rede de palavras-chave dos artigos levantados na revisão sistemática de satisfação no trabalho docente.

Figura 3 – Multigráfo da rede de palavras-chave.



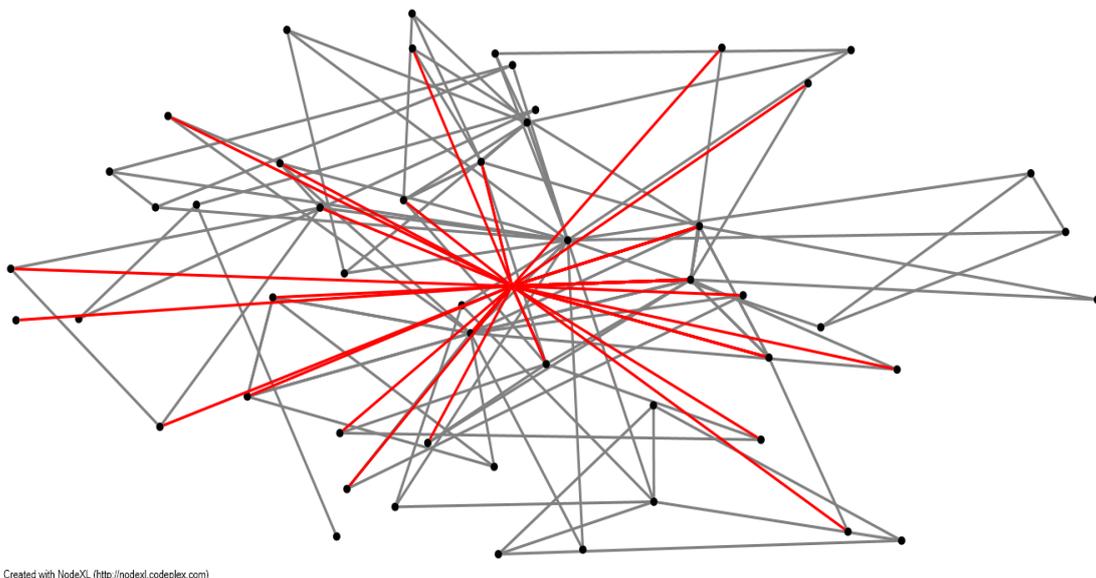
Fonte: Elaboração dos autores (2020).

A Figura 3 consiste de um multigráfo não direcionado, conexo, representando as associações entre todas as 52 diferentes palavras-chaves que são identificadas nos artigos sobre satisfação no trabalho docente. Cada ponto (vértice) representa uma palavra-chave e cada linha (aresta) representa as ligações entre as palavras-chave, ou seja, indica a frequência conjunta entre um par de palavras-chave nos artigos. Considera-se que o mesmo par de elementos (vértices) podem estar presentes em mais de um artigo. Neste multigráfo não são indicados os rótulos dos vértices e das arestas, pois o objetivo inicial é visualizar as

associações entre os termos de forma geral. Percebe-se, assim que, neste caso, não existem termos isolados, isto significa que as palavras-chave deste *dataset* estão associadas a outros descritores, indicando que a frequência conjunta entre os termos nos artigos atinge todas as palavras-chaves.

Um dos principais objetivos de se construir um grafo de relações é identificar quão próximos estão os elementos que o compõe. No entanto para explorar de forma mais aprofundada estas associações, pode-se identificar quais elementos são mais importantes. Para tanto, utilizou-se as métricas de centralidade, de intermediação e de relevância disponibilizadas pelo NodeXL. A Figura 4 apresenta uma das métricas utilizadas na Teoria dos Grafos, que indica o grau de importância de determinados vértices em relação aos outros elementos do multigráfo (centralidade Autovetor).

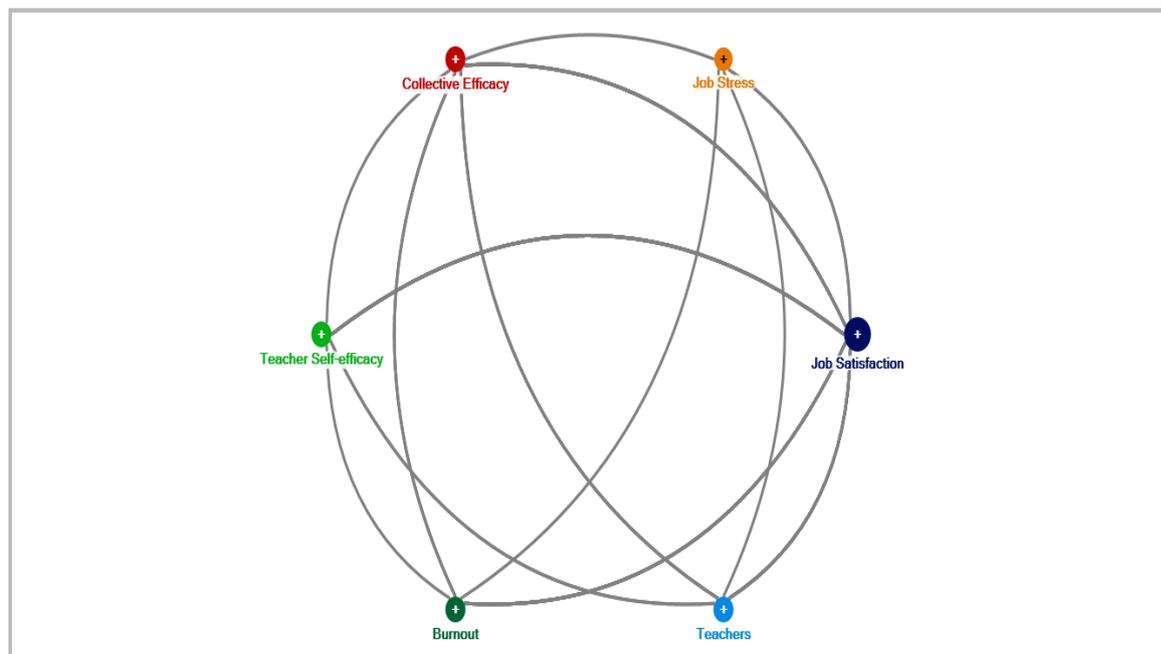
Figura 4 – Centralidade Autovetor das palavras-chave.



Fonte: Elaboração dos autores (2020).

Ao calcular a métrica de centralidade Autovetor (*Eigenvector Centrality*), pode-se identificar, neste caso, qual palavra-chave possui uma posição de maior importância no multigráfo. A métrica indica que o termo mais importante deste *dataset* é satisfação no trabalho, alcançando o maior escore (0,12). No multigráfo representado na Figura 4, a abrangência desta palavra-chave está destacada na cor vermelha. Percebe-se que a satisfação no trabalho é o termo mais relevante, indicando que sua frequência conjunta com outras palavras-chave nos artigos é a mais forte.

Após analisar de uma forma geral como está a composição das palavras-chaves, identificando suas associações, é possível realizar uma análise ainda mais detalhada, para visualizar as relações entre as principais palavras-chave do *dataset*. A Figura 5 representa as associações entre as principais palavras-chave.



Created with NodeXL (<http://nodexl.codeplex.com>)

Figura 5 - Multigráfico de associações entre as principais palavras-chave.

Fonte: Elaboração dos autores (2020).

A Figura 5 é um multigráfico compactado, não direcionado das principais palavras-chave dos artigos referentes à satisfação no trabalho. Para a construção deste multigráfico utiliza-se a métrica coeficiente de agrupamento (CG), que dividiu as palavras-chave em 6 grupos representados por seus termos mais importantes, a saber: grupo 1 - *satisfação no trabalho*; grupo 2 - *professor*; grupo 3 - *autoeficácia docente*; grupo 4 - *eficácia coletiva*; grupo 5 - *burnout*; grupo 6 - *estresse no trabalho*. Aplica-se o algoritmo *Clauset-Newman-Moore* para encontrar os grupos.

Consiste em um multigráfico pois houve a detecção de arestas paralelas duplas ou múltiplas entre diferentes pares de palavras-chave, tais como: *satisfação no trabalho* e *eficácia coletiva*; *professor* e *autoeficácia docente*; *estresse no trabalho* e *satisfação no trabalho*; entre outros. Ressalta-se que a quantidade de arestas entre um mesmo par de vértices só pode ser confirmada na planilha *Vértice* do NodeXL, não sendo possível visualizar

todas elas nos grafos ou multigráfos, pois se sobrepõe. Não foram identificados laços entre os termos, pois não houve relação de qualquer palavra-chave consigo mesma.

Para identificar os termos mais importantes de cada grupo, destacados na Figura 5, são calculadas as respectivas métricas: *degree*; *Betweenness Centrality*. *Closeness Centrality*, *Eigenvector Centrality*. As palavras-chave com maiores escores nessas métricas são consideradas as mais importantes de cada grupo, pois possuem maior capacidade de conectarem-se aos demais termos do cluster e maiores escores de relevância. Neste caso específico, não é calculada a métrica *PageRank*, pois este é um multigráfo não direcionado, a respectiva métrica deve ser aplicada apenas à grafos e multigráfos direcionados.

O referido multigráfo (Figura 5) possui subgrafos que formam cliques, pois os termos dos subconjuntos estão interconectados, tais como: satisfação no trabalho/eficácia coletiva/autoeficácia; burnout/estresse no trabalho/eficácia coletiva, entre outros. Os cliques formados nos subgrafos indicam que a coesão global é máxima (density, 1), desta forma, a conexão entre esses descritores é mais forte. Os dados foram estruturados para gerar uma representação geométrica compactada que representa o total de palavras-chave de cada grupo por uma figura específica, ressaltada a partir do seu termo mais importante. Enfatiza-se que os 6 clusters estão representados na Figura 5 por círculos. Quando maior for o círculo, mais importante é o grupo em relação a todos os demais do multigráfo. Assim o grupo 1, com 18 palavras-chave (vértices) e 58 ligações entre elas (arestas), representado pelo termo *satisfação no trabalho*, é o mais importante deste multigráfo, pois é o maior círculo da rede. Os demais grupos são ordenados, a seguir, por grau de importância: grupo 2 – *professor*, 13 palavras-chave e 25 ligações entre elas; grupo 3 – *autoeficácia docente*, com 8 descritores e 15 ligações entre eles; grupo 4 – *eficácia coletiva*, 6 palavras-chave e 9 ligações; grupo 5 - *burnout*, com 4 palavras-chave e 4 conexões; grupo 6 – *estresse no trabalho*, com 3 palavras-chave e 3 ligações entre elas. Para estabelecer o tamanho dos círculos representativos, o NodeXL utiliza as diferentes métricas de centralidade, de intermediação e de agrupamento anteriormente descritas. O grau de importância de um cluster em relação aos outros de qualquer grafo ou multigráfo é calculado levando em consideração a capacidade dos seus vértices de se conectarem a vértices de outros agrupamentos.

Considerando os dados analisados a partir de técnicas de análise com grafos no NodeXL, é possível identificar as regras de associação entre os artigos a partir das suas palavras-chave, identificando quais termos assumem a posição de intermediação e de centralidade entre os constructos explorados em estudos de revisão sistemática. Desta forma, tais técnicas de investigação e visualização das palavras-chaves podem servir para indicar ao

pesquisador quais construtos têm maior destaque no seu banco de dados, apontando possíveis tendências de investigação em determinada área do conhecimento, a partir dos artigos selecionados e referente ao período investigado.

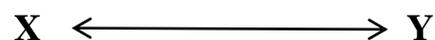
Identificando relações entre os resultados de revisão sistemática com o NodeXL

Para demonstrar a utilização de grafos em estudos de revisão sistemática, optou-se por investigar as conexões entre os resultados dos artigos sobre satisfação no trabalho docente. Explorou-se apenas as correlações positivas e negativas entre as variáveis dos estudos (Figura 6), pois são as relações mais analisadas pelos artigos investigados. Um delineamento correlacional (correlações positivas e negativas), permite concluir a forma da associação entre duas variáveis (magnitude e direção). Esta técnica estatística possibilita observar como as variáveis mudam, sem, no entanto, estabelecer uma relação causal. Indica-se que quando valores numa variável mudam, valores em outra também mudam de forma previsível (Dancey & Reidy, 2006)

Sendo assim para melhor compreender o multigráfo representado na Figura 7 é importante considerar que: (1) correlação positiva – cada vez que x aumenta, y aumenta um determinado valor de forma constante; (2) correlação negativa – cada vez que x aumenta, y diminui um valor de forma também constante. Deve-se considerar que quando uma seta é divergente de um vértice, isto significa que a relação parte deste vértice, por exemplo:

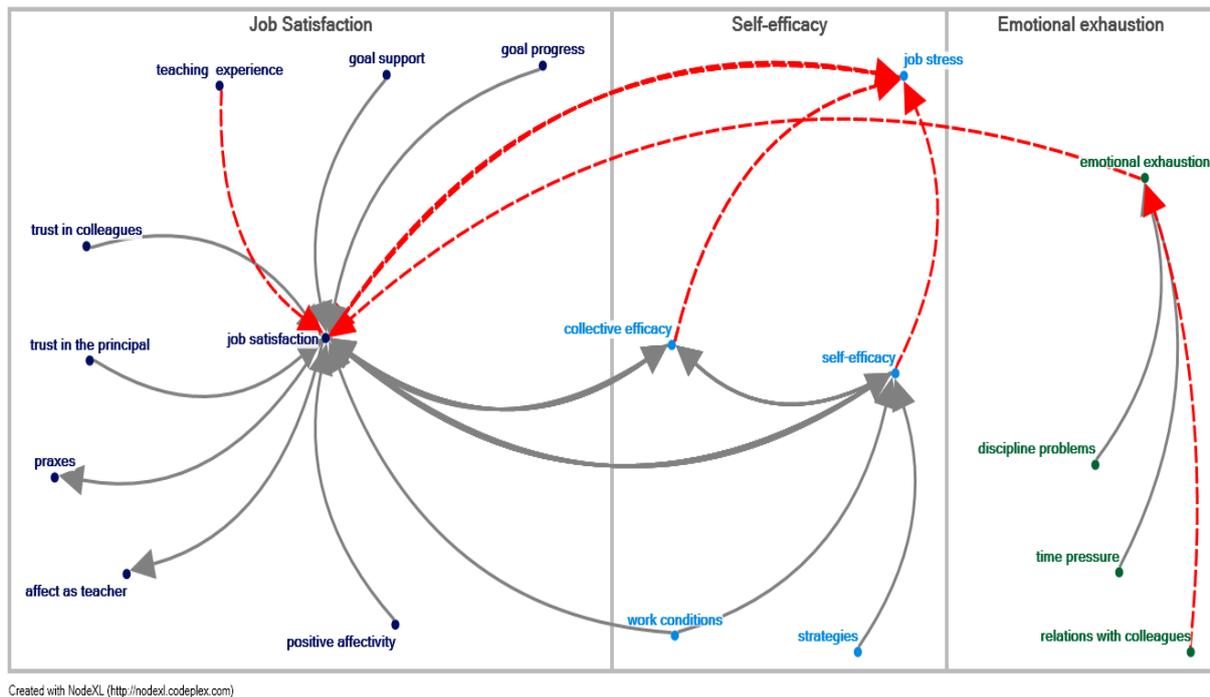


Como a seta é divergente de x, isto representa que, considerando uma possível correlação positiva, quanto mais o valor de x aumenta, mais o valor de y aumenta também. Se as arestas forem bidirecionais, isto representa que há associação mútua entre as variáveis



Enfatiza-se no caso de arestas duplas e/ou múltiplas entre um mesmo par de vértice, as setas são sobrepostas. As correlações positivas foram identificadas pelas arestas na cor cinza. As correlações negativas correspondem as arestas pontilhadas, na cor vermelha. Ressalta-se que o foco deste tópico não é analisar as relações entre as variáveis, mas sim explorar a utilização de grafos em revisão sistemática.

Figura 6 – Multigráfo direcionado das correlações entre as variáveis dos artigos.



Fonte: Elaboração dos autores (2020).

A Figura 6 consiste de um multigráfo direcionado expandido das correlações positivas e negativas entre as variáveis investigadas pela revisão sistemática de satisfação no trabalho, estruturado em clusters a partir da métrica Coeficiente de Agrupamento (CG), calculada com o algoritmo *Clauset-Newman-Moore*. Cada cluster é intitulado considerando seu vértice mais importante. O grau de importância dos vértices é calculado através das métricas *degree (in-degree, out-degree)*; *Betweenness Centrality*. *Closeness Centrality*, *Betweenness Centrality*, *PageRank*. Essas métricas indicam os graus de centralidade, intermediação e relevância dos vértices em grafos dirigidos. Assim os clusters são organizados por quadrante, agrupando-se os termos mais próximos, desta maneira formou-se 3 grupos, denominados: (1) *satisfação no trabalho*, (2) *autoeficácia*, (3) *exaustão emocional*. As setas na cor cinza correspondem as correlações positivas, setas pontilhadas e na cor vermelha indicam correlação negativa

Considera-se a Figura 6 como um multigráfo expandido, pois apresenta todos os vértices relacionados a cada grupo, não ocultando nenhum. É direcionado, pois indica uma direção para as relações entre as variáveis através de setas. Sendo assim, as arestas entre os vértices possuem uma direção e, portanto, o *degree* pode ser caracterizado como *in-degree*, indicando o número de arestas que os vértices recebem, e como *out-degree*, indicando o número de arestas emergentes do vértice.

Para analisar corretamente grafos direcionados é necessário entender a função e os posicionamentos das setas (arestas) que apresentam a direção da relação entre os vértices, identificando se a centralidade de grau é *in-degree* (\rightarrow) ou *out-degree* (\leftarrow). Ressalta-se que em casos de multigráfos, ocorrem arestas duplas ou múltiplas entre os vértices, quanto maior for a quantidade de ligações entre os vértices mais largas serão as arestas, indicando que a conexão entre os elementos envolvidos é forte. No entanto, a quantidade de ligações só pode ser exatamente visualizada nas colunas *in-degree* e *out-degree* disponíveis na planilha *Vértice*. Para gerar essas métricas é necessário escolher a opção *graph metrics* na barra de ferramentas do NodeXL e solicitar o cálculo de métricas específicas ou de todas as métricas da ferramenta. Automaticamente estes dados são calculados e apresentados na planilha *Vértice*. Ressalta-se que as arestas duplas ou múltiplas dos multigráfos podem seguir uma única direção (ou *in-degree* ou *out-degree*) tanto quanto podem seguir as duas direções (*in-degree* e *out-degree*), dependendo da natureza dos dados.

Na Figura 6, por exemplo, percebe-se, no grupo 1, que o progresso de objetivos está positivamente correlacionado com satisfação no trabalho, isto porque houve uma aresta de saída (*out-degree*) do vértice progresso de objetivos em direção ao vértice satisfação no trabalho. Ainda nesta mesma relação, nota-se que o vértice satisfação no trabalho recebe uma aresta, por isso configura como *in-degree*.

É possível avaliar a relação entre os vértices de grupos diferentes (grupos 1 e 2). As arestas que ligam os vértices satisfação no trabalho docente (grupo 1) e estresse no trabalho (grupo 2), por exemplo, formam setas mais largas em relação as demais dos outros vértices, o mesmo acontece com os pares de vértices satisfação no trabalho (grupo 1) e eficácia coletiva (grupo 2); satisfação no trabalho (grupo 1) e autoeficácia (grupo 2). Situações como estas indicam que a conexão entre estes elementos é mais forte que as outras, considerando todos os grupos deste multigráfo. Nota-se ainda, que nestes 3 casos em particular as setas seguem nas duas direções (*in-degree* e *out-degree*), sugerindo que estas variáveis se influenciam mutuamente, tanto com correlações positivas, no caso da satisfação no trabalho, da eficácia coletiva e da autoeficácia; quanto com correlações negativas, no caso da satisfação e do estresse no trabalho.

As métricas referentes aos clusters podem ser visualizadas nas planilhas *groups* e *groups vértices* do software NodeXL. É possível encontrar informações sobre a quantidades de vértices e arestas (únicas ou duplicadas) pertencentes à cada grupo, bem como valores máximos e mínimos de vértices e arestas conectados, o escore de coesão (*density*) de cada grupo, a média de distância, entre outras informações.

Tais técnicas de análise e visualização dos dados podem auxiliar pesquisadores de diferentes formas, dentre as quais destaca-se: (1) para representar graficamente os resultados dos artigos selecionados nas revisões sistemáticas favorecendo a construção de sínteses e a realização de comparação entre dados das pesquisas; (2) a organização dos dados por grupos (clusters) contribui para identificar variáveis que são investigadas em conjunto, tais informações que podem representar a forma como a literatura da área está estruturada teoricamente e indicando como alguns construtos-chave são investigados empiricamente; (3) as métricas auxiliam a detectar elementos mais importantes dentro de um banco de dados. Em relação aos resultados das pesquisas, tais informações podem servir para identificar variáveis já consolidadas em uma área de conhecimento e as que representam “zonas de fronteiras”, indicando variáveis não são tão fortemente investigadas pelos pesquisadores da área, mas que são relevantes pois sugerem um processo de renovação, com a inclusão de novos focos de análise; (4) de forma geral, as técnicas apresentadas podem auxiliar os pesquisadores a perceberem como os dados investigados se relacionam, indicando ainda a direção e a força (frequência da conexão) da relação.

Ressalta-se que escolheu-se apresentar como exemplo de aplicação dessas técnicas a análise das palavras-chave e dos resultados dos artigos selecionados para revisão sistemática, no entanto, pode-se fazer uso dessas técnicas para explorar outros tipos de dados, tais como os de caracterização dos artigos. Enfatiza-se ainda que decidiu-se apontar exemplos de análise dos resultados por meio de clusters para apresentar de forma mais ampla as possibilidades de investigação do programa computacional NodeXL, no entanto, é possível explorar os dados utilizando apenas as medidas dos vértices, sem usar métricas de grupo e/ou formação de representação gráfica organizada por cluster.

4. Considerações Finais

Este artigo metodológico explora o uso da ferramenta NodeXL para análise de dados e apresentação de resultados em estudos de revisão sistemática. Acredita-se que as informações disponibilizadas poderão auxiliar pesquisadores quanto à utilização de técnicas de análise de rede, a partir da Teoria dos grafos, para a exploração e visualização dos seus dados.

Neste sentido representa uma abordagem pioneira e eficaz de análise. Entende-se que a despeito de ser uma ferramenta mais direcionada para o estudo das relações das redes sociais, o NodeXL constitui-se um importante recurso para pesquisadores de qualquer área.

Assim, utilizando a ferramenta NodeXL, por meio de grafos, é possível identificar e visualizar regras de associação entre as palavras-chave dos artigos, bem como calcular e visualizar as relações entre variáveis de diferentes tipos, sendo muito eficaz na análise e apresentação dos dados. Neste estudo constatou-se que as técnicas de análise e visualização por meio de grafos, bem como a aplicação de medida de centralidade, demonstraram-se adequadas e eficazes, podendo ser utilizada em revisões sistemáticas.

Referências

Castro, A. A. (1992). Revisão Sistemática e Meta-análise (pp. 1-11).

Cordeiro, A. M., et al. (2007). Revisão sistemática: uma revisão narrativa. *Revista do Colégio Brasileiro de Cirurgiões*, 34(6), 428-431.

Dancey, C. P., & Reidy, J. (2006). Estatística sem Matemática para psicologia. Artmed.

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2011). *Analyzing Social Media Networks with NodeXL*. *Analyzing Social Media Networks with NodeXL* (pp. 11–29). Elsevier. <https://doi.org/10.1016/B978-0-12-382229-1.00002-3>

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2010a). *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. *Graduate Journal of Social Science* (Vol. 8, p. 304). Retrieved from <http://www.amazon.com/dp/0123822297>

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2010b). *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. (D. Hansen, B. Shneiderman, & M. A. Smith, Eds.) *Graduate Journal of Social Science* (Vol. 8, p. 304). Morgan Kaufmann. Retrieved from <http://www.amazon.com/dp/0123822297>

Locke, E. A. (1976). The nature and causes of job satisfaction. In *Handbook of Industrial and Organizational Psychology*, (pp. 1297–1349).

Matos, I. M. D. (2013). *Teoria dos grafos no ensino básico e secundário*. Universidade de Aveiro. Retrieved from <http://hdl.handle.net/10773/12083>

Pinheiro, J. M. V. (2013). *A investigação e as redes de conhecimento na European Network for Housing Research*. Porto: [Edição do Autor]. Retrieved from <http://repositorio-aberto.up.pt/handle/10216/72175>

Rother, E. T. (2007). Revisão narrativa vs revisão sistemática. *Acta Paulista de Enfermagem*, 20, 6–7.

Smith, M. A. (2013). NodeXL: Simple network analysis for social media. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 89–93. <https://doi.org/10.1109/CTS.2013.6567211>

Sousa, D. J. R. (2010, February 8). *Caracterização de ligações entre utilizadores em redes sociais*. Universidade do Porto. Retrieved from <http://repositorio-aberto.up.pt/handle/10216/57672>

Sousa, M. R., & Ribeiro, A. L. P. (2009). Revisão sistemática e meta-análise de estudos de diagnóstico e prognóstico: um tutorial. *Arquivos Brasileiros de Cardiologia*, 92(3), 241-251.

Porcentagem de contribuição de cada autor no manuscrito

Maély Ferreira Holanda Ramos– 100%

Fernando Augusto Ramos Pontes– 25%

Simone Souza da Costa Silva- 25%

Erika Cristina de Carvalho Silva Pereira- 25%