



Federação das Indústrias do Estado da Bahia

CENTRO UNIVERSITÁRIO SENAI - BAHIA

**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL E TECNOLOGIA INDUSTRIAL**

Dissertação de mestrado

**Redes semânticas baseada em cliques para apoio à
análise de conteúdo**

Apresentada por: Renata Souza Freitas Dantas Barreto
Orientador: Hernane Borges de Barros Pereira

Fevereiro de 2018

Renata Souza Freitas Dantas Barreto

Redes semânticas baseada em cliques para apoio à análise de conteúdo

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial, Curso de do Centro Universitário SENAI - BAHIA, como requisito parcial para a obtenção do título de **Mestre em Modelagem Computacional e Tecnologia Industrial**.

Área de conhecimento: Interdisciplinar

Orientador: Hernane Borges de Barros Pereira

Centro Universitário SENAI - BAHIA

Coorientador: Maria Teresinha Tamanini Andrade

IFBA

Salvador

Centro Universitário SENAI - BAHIA

Ano

Ficha catalográfica elaborada pela Biblioteca do Centro Universitário SENAI CIMATEC

B273r Barreto, Renata Souza Freitas Dantas

Redes semânticas baseadas em cliques para apoio à análise de conteúdo / Renata Souza Freitas Dantas Barreto. – Salvador, 2018.

63 f. : il., color.

Orientador: Prof. Dr. Hernane Borges de Barros Pereira.

Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) – Programa de Pós-Graduação, Centro Universitário SENAI CIMATEC, Salvador, 2018.
Inclui referências.

1. Redes semânticas - Cliques. 2. Modelo computacional. 3. Análise de Conteúdo. I. Centro Universitário SENAI CIMATEC. II. Pereira, Hernane Borges de Barros. III. Título.

CDD: 620.00113

Dedicado à Melquiades e Sarah.

Agradecimentos

Meus sinceros e mais profundos amáveis agradecimentos, Meu marido, melhor amigo e amor Melquiades.

Minha Filha, eternamente amiga e amada Sarah.

Minha Família, Marco, Márcia e Matheus Barreto.

Agradeço a Jair Tércio, Maribel Barreto, Paula Amorim (Fundação OCIDEMNTE).

Minhas amigas e meus amigos, Palmira, Camila, Cleônidas, Claudinei, Carina, Marcelo, Lamartine e Pat.

A minha querida banca, Professora Núbia Moura, Professor Marcelo Moret, Professor Eudaldo dos Santos Filho, Professora Teresinha Andrade.

Ao meu querido amigo “Brou” e orientador, Professor Hernane.

Desejamos que alguém desfrute desse trabalho tanto quanto nós. Esperamos do futuro a eterna e terna inspiração do presente que é a vida. Acreditando desse instante em diante que a força inteligente do universo, ou seja, Deus se manifesta através das Redes complexas. Pois, por mais que não percebamos ou ainda não tenhamos condições de revelar, as redes no mínimo são onipresentes.

Nos retiremos desejando: paz para todo universo, paz para todas as fronteiras, paz para toda humanidade terráquea, muita paz para toda o orbe terra. Que assim seja.

“Ora Lege Lege Lege Relege labora et invenies”. (14 prancha Mutus Liber, 1677).

Salvador, Brasil
dia de Fevereiro de 2018

Renata Souza Freitas Dantas Barreto

Resumo

Este estudo versa sobre Redes Semânticas e Análise de Conteúdo (AC) considerando às suas similaridades e implicações. Pretendemos compreender e demonstrar como as redes semânticas baseadas em cliques podem subsidiar o método da AC que demanda muito tempo e labor, independente da extensão do discurso analisado. O objetivo desta pesquisa é elaborar um modelo computacional que auxilie o processo de identificação dos Núcleos de Sentidos (NS), Categorias e conferências das Unidades de Contexto (UC) na AC. O nosso modelo se baseia no uso de algumas propriedades de redes, especificamente grau médio, densidade, modularidade e, conseqüentemente, comunidades. Usamos como fontes de dados (1) os discursos e a AC de uma pesquisa sobre Colaboração científica e Difusão do Conhecimento relativas às redes de coparticipação de pesquisadores em um programa de pós graduação e (2) os relatórios técnicos de educação (Educação para Todos - EPT) da Organização das Nações Unidas para Educação, Ciência e Cultura (UNESCO). Testamos o modelo com a primeira fonte de dados, a fim de verificar a resposta do modelo; em seguida, aplicamos o modelo à segunda fonte de dados e analisamos os resultados considerando as exigências da AC. Os resultados indicam que é possível, com a análise de redes semânticas baseadas em cliques, tornar célere e aumentar a acurácia da criação dos núcleos de sentidos, categorias e análise das unidades de contexto na AC.

Palavras-chave: Redes semântica, Cliques, Modelo computacional, Análise de Conteúdo.

Abstract

This study deals with Semantic Networks and Content Analysis (CA) considering its similarities and aids. We intend to understand and demonstrate how semantic networks based on clicks can subsidize the CA method that demands a lot of time and labor, regardless of the extent of the discourse analyzed. The objective of this research is to elaborate a computational model that optimizes the process of identification of the Sense Cores (NS), Categories and Conferences of the Contextual Units (UC) in the AC. Our model is based on the use of some properties of networks, specifically medium grade, density, modularity and, consequently, communities. We used as sources of data (1) the discourses and the AC of a research on "Scientific Collaboration and Diffusion of Knowledge" concerning the networks of participation of researchers in a graduate program and (2) the technical reports of education All - EPT) of the United Nations Educational, Scientific and Cultural Organization (UNESCO). We tested the model with the first data source in order to verify the model response; then we apply the model to the second data source and analyze the results considering the requirements of the CA. The results indicate that it is possible, with the analysis of semantic networks based on clicks, to accelerate and increase the accuracy of the creation of sense nuclei, categories and analysis of context units in the CA.

Keywords: Semantic Networks, Cliques, Computational Model, Content Analysis.

Sumário

1	Introdução	1
1.1	Definição do problema	2
1.2	Objetivo	2
1.3	Importância da pesquisa	3
1.4	Aspectos metodológicos	3
1.5	Organização da dissertação	4
2	Bases Teóricas: Redes Semânticas e Análise de Conteúdo	5
2.1	Teoria de Redes Complexas	5
2.1.1	Principais propriedades de redes e seus significados	6
2.2	Topologia das Redes	9
2.2.1	Definições Topológicas das Redes Complexas	10
2.3	Redes Semânticas	11
2.3.1	Redes semânticas baseadas em cliques (RSBC)	12
2.4	Comunidades	14
2.4.1	Modularidade	16
2.4.2	Algoritmo de Louvain	18
2.5	Análise de Conteúdo	20
2.5.1	Principais características da Análise de Conteúdo	21
3	Modelo Computacional Proposto: redes semânticas de cliques	24
3.1	Método da Pesquisa: Modelo Computacional Proposto	24
3.1.1	Distribuição espacial das redes semânticas de cliques em comunidades (DERC)	26
3.2	Fonte de dados, os seus cenários e os instrumentos utilizados	30
3.2.1	Cenário 1: tese doutoral	30
3.2.2	Cenário 2: Relatório de Monitoramento Global: EPT 2015	31
4	Resultados e Discussões	33
4.1	Propriedades convergentes na Análise de Conteúdo e das Redes Semântica de Cliques	33
4.2	Aplicação do Modelo Computacional Proposto: Análise experimental/Simulações	35
4.2.1	Simulação no Cenário 1	35
4.2.2	Simulação no Cenário 2	38
4.2.3	Simulação segundo cenário	38
4.3	Resultados cenário 1	41
4.4	Resultados cenário 2	44
5	Considerações Finais e Contribuições	47
5.1	Trabalhos futuros	48
	Referências	50

Lista de Tabelas

3.1	Categorias e Núcleos de Sentido do cenário 1.	30
3.2	10 palavras com maior frequência do NS1 e NS2.	31
4.1	Comparação das 10 palavras de maior frequência do Núcleo de Sentido 1 (NS1: Difusão do Conhecimento - Categoria 1: Difusão do Conhecimento e Colaboração) com seu grau na Rede.	42
4.2	Comparação das 10 palavras de maior frequência do Núcleo de Sentido 5 (NS5: Comunidade Acadêmica - Categoria: Estrutura Acadêmica) com seu grau na Rede.	43
4.3	As 10 palavras que compõem o Núcleo de Sentido 7 (NS7: Política de Publicação - Categoria 3: Política de CTI)	43
4.4	Tabela das CAT e NS a partir da aplicação do nosso modelo computacional.	45

Lista de Figuras

2.1	Rede alimentar de interações predador presa	6
2.2	Cálculo do grau de um vértice e os elementos que compõem a rede.	7
2.3	Rede Aleatória e rede Livre de Escala	8
2.4	Exemplo de uma rede mundo pequeno.	10
2.5	Rede semântica de palavras que compõem um Núcleo de Sentido da AC de um estudo sobre “Difusão do conhecimento”.	12
2.6	Formação das redes de cliques por justaposição e superposição.	13
2.7	Formação das redes de cliques por justaposição.	13
2.8	Rede de telefonia móvel belga	14
2.9	Definições modularidades	18
2.10	Etapas do procedimento do algoritmo de Louvain	20
3.1	Fluxograma da metodologia do Modelo computacional	25
3.2	Fluxograma da metodologia do DERC	27
3.3	Exemplo das 3 primeiras Etapas da metodologia do DERC	28
3.4	Essa imagem representa todo o processo da DERC.	29
3.5	Capas dos EPT já realizados.	32
4.1	Uma rede semântica composta por duas cliques por justaposição	34
4.2	Fluxograma da aplicação do MC proposto no primeiro cenário da amostragem	36
4.3	Rede semântica de todos os discursos do cenário 1	37
4.4	Modelo computacional aplicado no segundo cenário.	38
4.5	Rede geral com as 7 comunidades e com as suas ligações.	40
4.6	Rede geral com as 7 comunidades e sem as suas ligações.	41
4.7	Sentido na elaboração da autora Andrade: Categorias, Núcleos de Sentido e Unidade de Registro da autora Andrade (2013)	42
4.8	Sentido da nossa elaboração: Unidades de Registro, Núcleos de Sentido e Categoria da nossa pesquisa	42
4.9	Comunidade 2 gerada pelo algoritmo de modularidade com suas 20 palavras com maiores vértices.	44
4.10	Comunidade 6 gerada pelo algoritmo de modularidade com suas 20 palavras com maiores vértices.	45
5.1	Rede dos processados, investigados ou réus da Lava Jato.	48
5.2	Rede Semântica das entrevistados da NA	49

Lista de Siglas

PPGMCTI ..	Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial
AC	Análise de Conteúdo
C1	Categoria 1
C2	Categoria 2
CAT	Categoria
DERC	Distribuição espacial das redes semânticas de cliques em comunidades
EPT	Educação para todos
MC	Modelo computacional
NS	Núcleo de Sentido
RS	Rede Semântica
RSBC	Rede Semântica baseada em clique
UC	Unidade de Contexto
UR	Unidade de Registro

Introdução

A pesquisa sobre redes semânticas advém de um estudo mais específico sobre a teoria de redes complexas, uma ciência nova, que carrega em sua estrutura básica a interdisciplinaridade. A cada ano de amadurecimento científico desta ciência, novas propriedades de redes surgem na expectativa de melhor caracterizá-la e sua aplicabilidade aumenta na medida que os pesquisadores, das mais diversificadas áreas do conhecimento, apropriam-se desse novo campo de investigação.

Redes semânticas são uma forma de representar o conhecimento extraído de um discurso escrito ou oral, sendo visualizado através de grafos. Os vértices de um grafo G , compõem o conjunto denotado por $V(G)$, e seus elementos são denominados de vértices ou nós.

As suas conexões são chamadas de arestas é descrita por $E(G)$. Matematicamente os grafos são descritos da seguinte forma: $G = (V, E)$ que consiste em dois conjuntos V (finito e não vazio) e E (relação binária sobre V).

Esta dissertação surge quando percebemos a necessidade de elaborar um modelo computacional que responda as perguntas para da nossa pesquisa, essas que não conseguem ser respondidas somente com abordagem metodológica qualitativa.

Ao percebermos que a Análise de Conteúdo proposta por [Bardin \(1977\)](#) possui limitantes por em muitos momentos ser demaseadamente árduo o trabalho manual em documentos extensos, é assertivo falarmos que um Modelo Computacional de Redes Semânticas baseado em cliques poderá contribuir otimizando os procedimentos da AC.

Segundo [Bardin \(2010\)](#), AC é um conjunto de técnicas para analisar discursos buscando obter, através de procedimentos sistemáticos, indicadores (quantitativos ou não) que possibilitem a conclusão de conhecimentos sobre a origem destes discursos.

Na execução da análise de conteúdo há uma necessidade de realizar contagem de palavras existente no discurso, a fim de contabilizar as suas frequências e em seguida verificar os contextos onde estão inseridas no texto. Após esse primeiro momento, o pesquisador destrincha o discurso, analisando as palavras denominadas de Unidade de Registro (UR) para posteriormente agrupá-los em Categorias (CAT). Autores como [Cunha \(2012\)](#) e [Andrade \(2013\)](#) se utilizaram de mais um procedimento para facilitar a análise e Categorização das UR, a elaboração de subgrupos denominado de Núcleos de Sentido (NS). Todos esses processos são exaustivos e demanda muito tempo.

Desse modo, para propormos um modelo computacional que aprimore o processo de detecção dos NS, Categorias e Unidade de contexto (UC) na Análise de Conteúdo, inicialmente, necessitaremos abordar e aprofundar o estudo em redes semânticas de cliques proposto por [Fadigas e Pereira \(2013\)](#); e nos estudos das propriedades das redes complexas.

Este trabalho tem como foco principal a elaboração do modelo computacional supracitado e a averiguação do seu possível auxílio na AC.

1.1 Definição do problema

O procedimento da Análise de Conteúdo de Bardin independentemente do número de páginas do discurso escolhido é demasiadamente trabalhoso de ser executado. Se não conseguirmos realizar a AC de maneira mais célere e menos laborosa, muito difícil será utilizar esse método de análise em documentos extensos, ao ponto de ser quase inviável realizar, gerando problemas na análise.

1.2 Objetivo

O objetivo geral deste trabalho consiste em compreender e demonstrar como as RS de cliques são capazes de auxiliar no processo de execução do método da AC. Através da elaboração de um modelo computacional (MC) capaz de auxiliar na identificação das Unidades de registro, Unidades de contextos, Núcleos de sentido, Categorias para a realização da Análise de Conteúdo.

Já para os objetivos específicos, analisaremos a relação entre algumas propriedades de redes semânticas, principalmente os relacionados a importância de vértices (i.e. grau médio, densidade e modularidade) e a estrutura do método de Análise de Conteúdo.

A importância do estudo referente as propriedades de redes e estudo dos procedimentos do método da AC, é essencial pelo fato de só assim ser possível para articular uma análise comparativa entre os índices de rede e os resultados da análise de conteúdo.

Em tempo, analisaremos as redes semânticas a partir de duas fontes de dados (1) tese doutoral e (2) relatório técnico sobre educação.

1.3 Importância da pesquisa

Quando analisamos um documento utilizando uma rede semântica, podemos identificar, de maneira assertiva, a relevância e os significados de uma palavra inserida em um texto, além de ser possível perceber os contextos nos quais as palavras estão imersas. Os discursos relevantes são aqueles que estão frequentemente envolvidos na relação com outros discursos e este envolvimento os torna mais relevantes na rede.

O método da AC que é um processo desgastante de ser realizado, ao elaborarmos um modelo computacional de RS baseado em cliques, pretendemos otimizar o processo de identificação dos Núcleos de Sentidos, Categorias e conferências das Unidades de Contexto na AC. Sendo assim, a viabilidade de realizar a AC em documentos robustos seria possível, além da redução de tempo e esforço do pesquisador, em comparação ao método original de [Bardin \(2010\)](#).

1.4 Aspectos metodológicos

Optamos em realizar na nossa pesquisa uma abordagem de técnicas quantitativas integradas às qualitativas, assim poderemos aproveitar o melhor que cada uma tem à oferecer no âmbito da pesquisa científica.

A pesquisa quantitativa é objetiva, visa a precisão das informações. Prioriza em sua análise as generalizações de padrões, noções e categorias, sendo estas replicáveis em diversas situações. Por sua vez, a pesquisa qualitativa tem como objetivo principal o contexto, o subjetivo. Segundo [Goldenberg \(2001\)](#), a pesquisa qualitativa é uma compreensão profunda, priorizando muito mais as descrições e entendimentos do universo estudado às generalizações numéricas.

Os dados foram extraídos de duas fontes distintas: (1) os documentos de uma pesquisa doutoral e (2) relatórios técnicos de educação da UNESCO.

Para realizar este trabalho, utilizaremos um método de construção de redes semânticas que consiste basicamente em (1) eliminação das palavras sem significados relevantes e conectivos (2) e na alteração das palavras restantes para sua forma canônica.

A rede semântica será formada por cliques ou subgrafos completos, utilizaremos o algoritmo de modularidade para dessa forma, poderemos estudar as comunidades estabelecidas na rede semântica e conferir a relevância dessas cliques para o método da AC.

Para a elaboração e estudo sobre redes semânticas teremos como base as dissertações de mestrado de [Caldeira \(2005\)](#) e [Teixeira \(2007\)](#); o estudo mais específico sobre Rede semânticas de Cliques os artigos de [Fadigas et al. \(2009\)](#) e [Pereira et al. \(2011\)](#) e para os estudos sobre modularidade e comunidade, respectivamente, o artigo de [Newman \(2003\)](#), o artigo de [B. Jean-L. G. e L. \(2008\)](#) e o livro de [Barabási \(2016\)](#).

O nosso modelo computacional teve enquanto suporte teórico e metodológico os trabalhos supracitados, que serão devidamente detalhados no Capítulo 3. A escolha dos referenciais teóricos levou em consideração as aplicabilidades e eficiências das RS de cliques no nosso caso de estudo.

1.5 Organização da dissertação

Este documento apresenta 5 capítulos e está estruturado da seguinte forma:

- **Capítulo 1 - Introdução:** Contextualização sobre os principais pontos a serem trabalhados nesta pesquisa.
- **Capítulo 2 - Bases Teóricas: Redes Semânticas e Análise de Conteúdo:** Capítulo responsável por apresentar as propriedades e conceitos sobre a teoria de redes, redes semânticas baseada em cliques e Análise de conteúdo e suas características;
- **Capítulo 3 - Modelo Computacional Proposto:** Capítulo responsável por apresentar o modelo proposto e sua aplicação nas fontes de dados;
- **Capítulo 4 - Resultados e Discussões:** Capítulo responsável por apresentar os resultados obtidos, discussão e explanação do modelo computacional proposto.
- **Capítulo 5 - Considerações Finais:** responsável por apresentar as considerações finais, bem como as contribuições da pesquisa e atividades futuras a serem desenvolvidas a partir dos resultados obtidos nesta dissertação.

Bases Teóricas: Redes Semânticas e Análise de Conteúdo

Neste capítulo desenvolveremos aspectos relacionados ao conceito e às propriedades da teoria de redes complexas com o enfoque nas Redes Semânticas, principalmente, as baseadas em cliques. Além de abordarmos os fundamentos epistemológicos da análise de conteúdo de Bardin.

2.1 Teoria de Redes Complexas

As pesquisas sobre grafos ou redes, sejam complexas e/ou sociais, possuem bastante relevância no meio científico, dada a capacidade de uma rede poder representar através de modelagens percepções aproximadas da natureza, sendo utilizadas para solucionar diversas categorias de problemas.

O termo redes complexas refere-se a um sistema que apresenta uma estrutura e propriedades descritivas relevantes, composto por um conjunto de vértices (ou nós) que são conectados por meio de arestas [Barabási \(2009\)](#). É importante salientar que nem toda rede é considerada uma rede complexa, pois essa classificação só é possível se o grafo apresentar propriedades topológicas que não encontramos em grafos simples.

Uma diversidade de autores acredita que o artigo publicado por Leonhard Euler (1707-1783), datado do ano de 1736, denominado de *Solutio problematis ad geometriam situs pertinentes*, demarcou a origem da teoria dos grafos [Euler \(1736\)](#). O pesquisador Leonhard apresentou a resposta para o problema do passeio nas pontes de Königsberg (atualmente conhecida por Kaliningrado), demonstrando que era impossível um seguir em uma rede que tenha mais de dois nós (vértices) com um número ímpar de linhas (arestas).

Uma rede pode ser apresentada matematicamente por $G = (V, E)$, onde o conjunto V é finito e não vazio e E relação binária sobre V . Os elementos de V são denominados de vértices e os elementos de E são denominados de arestas [Gross e Yellen \(1999\)](#). Seguindo essa perspectiva, uma rede permite o estudo e análise desses elementos, denominados de nós ou vértices e suas interações denominadas de arestas.

Quando o sistema possui muitos vértices, com n tendendo ao infinito, podemos observar

propriedades de rede emergirem. O surgimento das propriedades é basilar para ela entrar na categoria de redes complexas. A Figura 2.1 é uma rede complexa, que representa as interações entre predador-presa de um determinado meio ambiente.

Na Subseção 2.1.1 veremos parâmetros (*e.g.* grau, densidade) de redes com mais detalhes e na Seção 2.2 descrevemos as topologias mais conhecidas de redes (i.e. aleatória; livre de escala e rede mundo pequeno). No entanto, antes de abordarmos sobre as topologias das redes é necessário conhecermos algumas propriedades que as definem.

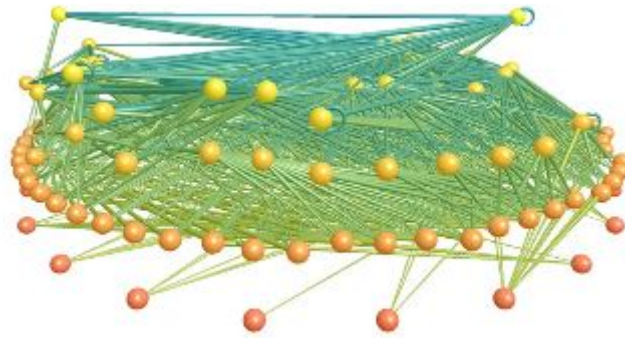


Figura 2.1: Rede alimentar de interações predador presa entre espécies em um lago de água doce. Fonte: Newman (2003, p.170).

Por fim, é importante frisar que as redes analisadas em nossa pesquisa são não dirigidas, ou seja, as arestas não possuem uma direção de saída e entrada para conectar os vértices.

2.1.1 Principais propriedades de redes e seus significados

As redes complexas apresentam propriedades que são imprescindíveis para sua análise e classificação. Há propriedades globais (que abrangem toda a rede) e propriedades locais (dizem respeito ao vértice). Abaixo são apresentadas algumas das principais propriedades de redes que nos últimos anos têm recebido atenção na literatura e seus respectivos conceitos.

Os **Vértices** - $V(\mathbf{G})$ - são os elementos que compõem a rede, denota por n . Já as **Arestas** ($E(\mathbf{G})$) são os segmentos de linhas que representam as conexões entre os vértices, referidas também por (m) .

O **Grau** (k) representa o número de conexões do vértice. Sendo o Grau o número de ligações de cada vértice, o **Grau Médio** é à média de conexões realizada por um vértice. A Figura 2.2 traz um exemplo simplificado sobre o cálculo do grau de um vértice.

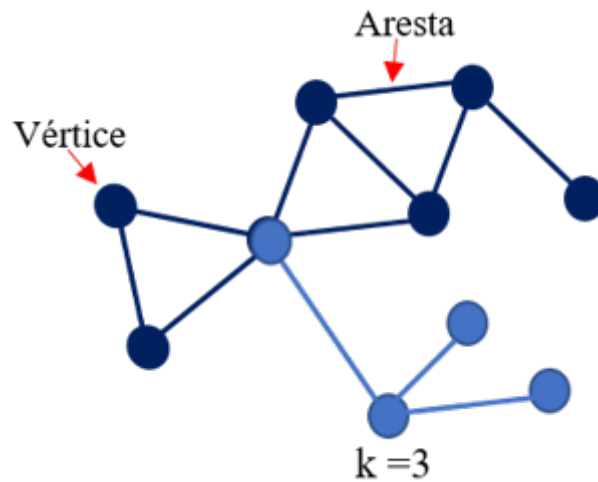


Figura 2.2: Cálculo do grau de um vértice e os elementos que compõem a rede. Fonte: Autora.

A **Distribuição de Grau** - $P(k)$ - representa o histograma dos graus dos vértices. A distribuição de probabilidade do número de conexões de todos os vértices em toda a rede. A distribuição de grau é uma das propriedades que permitem a caracterização topológica das redes, por conta de demonstrar a forma que os vértices estão conectados entre si [Newman \(2003\)](#).

Como, por exemplo, se uma rede topologicamente for definida como **Livre de Escala**, isso significa que as conexões entre os seus vértices seguem uma Lei de potência (i.e. poucos vértices com muitas conexões - *Hubs*). Por outra via, se a rede for caracterizada como **Aleatória**, as ligações entre seus vértices seguem uma *Poisson*.

[Barabási \(2016\)](#) ilustra na Figura 2.3 a calda da distribuição de *Grau* das redes, comparando a Lei de potência (Livre de Escala) e Poisson (Aleatória).

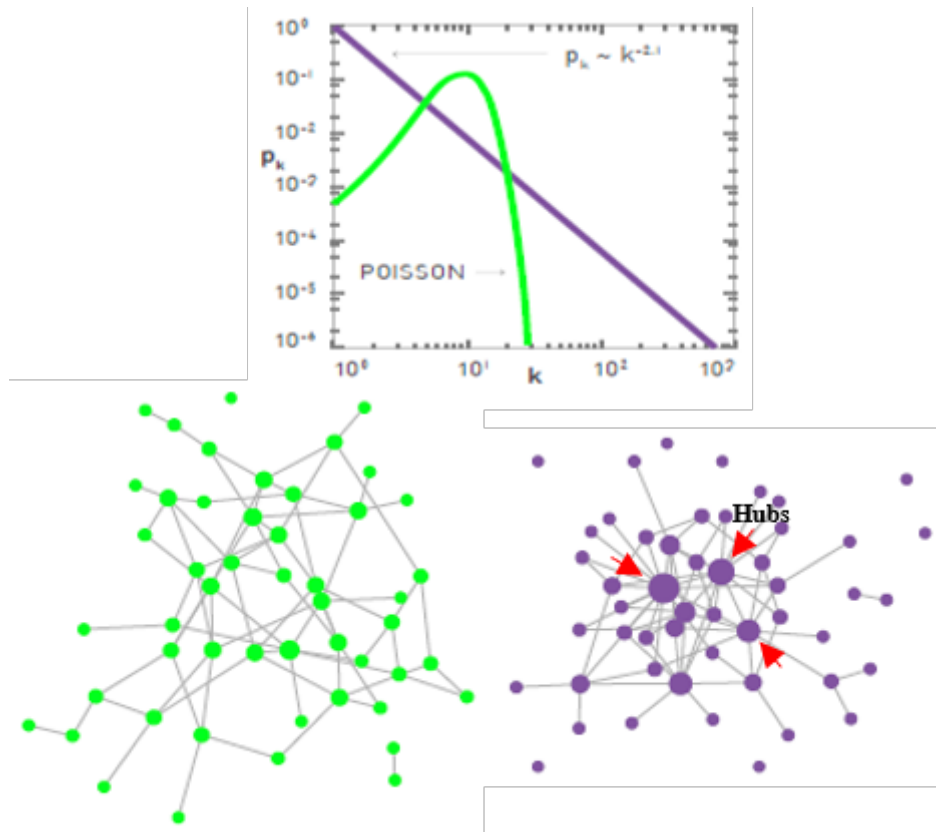


Figura 2.3: A rede com os vértices pintados de verde representa uma rede Aleatória. A rede com os vértices pintados de roxos representa uma Livre de Escala. O quadro em *loglog* é a representação matemática das distribuições de graus das redes apresentadas. Fonte: [Barabási \(2016\)](#).

O conceito de **Densidade** (Δ) para a teoria de redes é a razão entre o número existente de arestas e o número máximo de arestas possíveis nessa mesma rede. A Densidade de uma rede não dirigida é calculada pela Equação:

$$\Delta = \frac{2E}{n(n-1)} \tag{2.1}$$

O **Caminho mínimo médio (CMM)** - (L) é a média das distâncias das geodésicas (menor distância que une dois vértices), ou seja, é a média de todos os caminhos mínimos dos vértices. O CMM nos auxilia para dizermos se a rede é eficiente no fluxo de informação ou não. O CMM entre os pares de vértice é dada pela Equação:

$$l = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij} \tag{2.2}$$

onde, d_{ij} é a distância geodésica do vértice i ao vértice j .

Assim, o **maior CMM** é chamado de **Diâmetro** (d) da rede, ou seja, é a maior menor distância entre dois vértices. Ao descobrirmos o diâmetro da rede, compreendemos qual a maior distância que uma informação terá que percorrer entre dois vértices.

Coefficiente de aglomeração (CA) é conceituado como as conexões entre o número de arestas existentes entre os vértices vizinhos do vértice analisado e o número máximo possível de arestas de sua vizinhança. O seu cálculo é obtido pela Equação:

$$C_i = \frac{2 \times E_i}{k_i \times (k_i - 1)} \quad (2.3)$$

onde C_i representa o CA do vértice i . E_i é o número de arestas existentes entre os vizinhos do vértice i . O k_i é o grau do vértice analisado, e $k_i \times (k_i - \frac{1}{2})$ representa o número máximo de arestas entre os vizinhos do vértice i .

O CA varia entre 0 à 1. Quanto mais próximo de 1 for o CA do vértice, mais os seus vizinhos tendem a formar triângulos. Sendo assim, o **Coefficiente de aglomeração médio - CAM** (C) - é a média dos CA de todos os vértices da rede.

Há também propriedades importantes, especificamente, relacionadas as medidas de centralidade do vértice. Com a disseminação das pesquisas em redes, os estudiosos têm introduzido novas medidas de centralidade com o objetivo de medir a variação da relevância dos vértices em suas redes.

Desta maneira, a centralidade de grau é a medida da influência direta que um vértice tem em relação a seus vizinhos, a proximidade está relacionada com a duração que uma certa informação leva para ser propagada por todos os vértices na rede, e a intermediação de um vértice está relacionada com as possibilidades de que esse vértice esteja no menor caminho entre outros vários pares de vértices.

Para aprofundar os estudos nas propriedades de redes, aconselhamos a leitura dos seguintes autores, onde buscamos referência para a síntese supracitada: [Newman \(2003\)](#), [Watts \(1999\)](#) e [Barabási \(2016\)](#).

2.2 Topologia das Redes

Consideramos três topologias de redes que seguem: rede de mundo pequeno, rede livre de escala e rede aleatória. A caracterização da rede em uma dessas três topologias advém dos cálculos das propriedades das redes.

2.2.1 Definições Topológicas das Redes Complexas

Redes aleatórias: Nesse modelo, arestas não direcionadas são acrescentadas de forma aleatória em um número fixo de vértices. As arestas são representadas com base em alguma probabilidade p . Os graus de cada vértice na rede segue a distribuição de Poisson, tendo como limite máximo N .

Uma maneira simples de identificar uma rede aleatória é realizar a distribuição de grau dos vértices da rede (histograma) e analisar se sua curva segue a distribuição de Poisson. Essa distribuição indica que a maioria dos vértices possuem o número de ligações próximo de um valor médio.

Pelo importante papel desempenhado pelos matemáticos Pául Erdős e Alfréd Rènnyi, na compreensão das propriedades dessas redes, em sua homenagem, as redes aleatórias são denominadas de rede Erdős-Rènnyi (BARABÁSI, 2016).

Rede de mundo pequeno: Watts e Strogataz (1998) acreditavam que muitas redes apresentam padrões altamente conectados. A Figura 2.4 ilustra uma rede mundo pequeno.

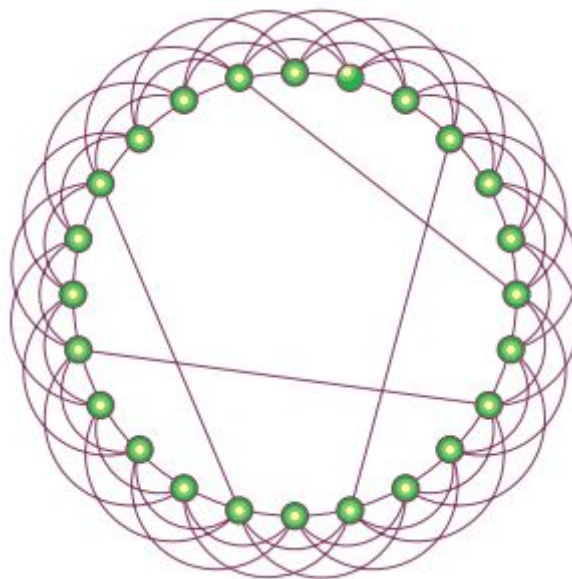


Figura 2.4: Exemplo de uma rede mundo pequeno. Fonte: Strogatz (2001, p.271)

As redes mundo pequeno possuem duas características básicas: o coeficiente de aglomeração alto e caminho mínimo médio baixo. Esse tipo de rede, dentre suas utilidades, proporciona, por exemplo, que a informação (e.g. vírus, notícias etc.) se propague de forma rápida e eficiente por ter justamente o CA alto e o CMM baixo.

Para o aprofundamento sobre as redes mundo pequeno e suas regras de restrições, aconselhamos a leitura do trabalho de [Watts e Strogatz \(1998\)](#).

Rede livre de escala: Os autores [Albert e Barabási \(2000\)](#) são os responsáveis por propor um modelo de configuração para redes livres de escala. “A principal diferença entre uma rede aleatória e uma rede sem escala vem na cauda da distribuição de grau”¹ [Barabási \(2016\)](#). A rede se constitui a partir do princípio da ligação preferencial.

Uma das características fundamentais das redes livres de escala é o fato de sua distribuição de grau seguir uma lei de potência. Isto ocorre porque no instante da criação e expansão da rede, as conexões entre os vértices surgem em função do grau dos vértices que estão sendo conectados (formando *Hubs*).

O resultado desse modelo termina por fazer que poucos vértices possuam muitas conexões (um grau muito alto), enquanto a grande massa de vértices se conectem pouco e possuam um grau baixo. Em redes livres de escala há uma tendência a prevalecer uma dinâmica de estruturação baseada na importância do vértice, em um padrão que foi nomeado por [Barabási \(2009\)](#) como os “ricos ficam mais ricos”. A [Figura 2.3](#) retrata a rede livre de escala, com os vértices pintados da cor roxa.

2.3 Redes Semânticas

Uma rede semântica (RS) é um conjunto de elementos conectados que compõem uma rede utilizando palavras, entidades ou conceitos, que se encontram relacionados pelos seus significados [Sternberg \(2011\)](#). [Grilo et al. \(2017, p.95\)](#) definem RS como “um sistema para representar o conhecimento imbuído de intenção de funcionalidade e estabelecido pelo contexto. Uma rede semântica é composta por palavras, conceitos ou entidades com significado semântico e seus relacionamentos.”²

Os vértices representam palavras dos discursos orais, ou de textos escritos (i.e. entrevistas, títulos de artigos, teses) e as arestas, a relação de coexistência das palavras em uma mesma sentença ou título. Em teoria das redes, há possibilidade de descrevermos e evidenciarmos como uma rede de conceitos se manifesta e evolui [Caldeira \(2005\)](#) e [Teixeira \(2007\)](#).

As redes semânticas demonstram como o conhecimento humano é estruturado através dos conceitos evocados em discursos falados e/ou escritos, a partir de diferentes estímulos ou

¹The main difference between a random and a scale-free network comes in the tail of the degree distribution ([BARABÁSI, 2016](#))

²a semantic network is a system for representing knowledge imbued with an intention of functionality and established by context. A semantic network is composed of words, concepts or entities with semantic meaning and their relationships. [Grilo et al. \(2017, p.95\)](#).

perspectivas. A técnica da modelagem de um texto ou de um discurso por meio de RS fornece um meio empírico de acesso à estrutura mental do conhecimento de quem proferiu o discurso [Sternberg \(2011\)](#). A Figura 2.5 é um exemplo de Redes de palavras:

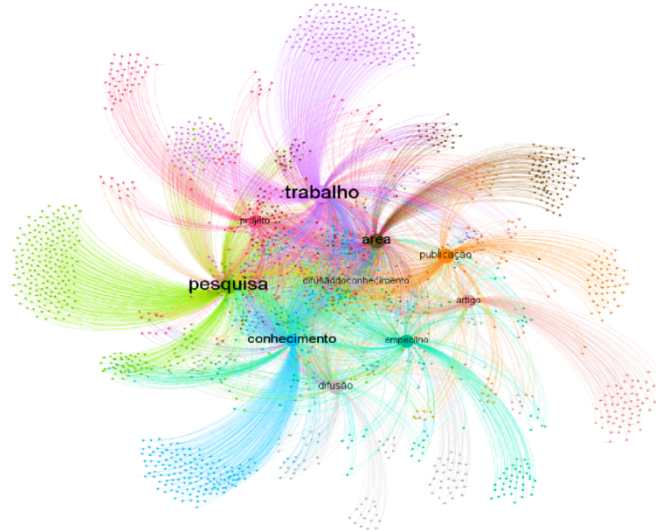


Figura 2.5: Rede semântica de palavras que compõem um Núcleo de Sentido da AC de um estudo sobre “Difusão do conhecimento”. Fonte: [Barreto et al. \(2016\)](#).

Analisar um conceito utilizando uma rede semântica, nos proporciona identificar, de maneira assertiva, a relevância e os significados de uma palavra inserido em um texto, levando ainda em consideração ser possível perceber os contextos nos quais as palavras estão imersas.

2.3.1 Redes semânticas baseadas em cliques (RSBC)

O pesquisador [Barabási \(2016\)](#) define as cliques como subgrafos completos ³, ou seja, onde todos os seus vértices estão conectados entre si. Este comportamento constitui-se em uma ocorrência comum em redes semânticas, representando sentenças ou títulos, onde cada palavra está conectada com todas as outras. Essa tendência inerente à aglomeração é quantificada pelo **Coefficiente de Aglomeração Watts e Strogatz (1998)** apud [Albert e Barabási \(2000\)](#).

Segundo [Fadigas e Pereira \(2013\)](#) as redes formadas por cliques baseiam-se na concepção de que as cliques inicialmente isoladas se conectam por justaposição e/ou superposição. Ainda segundo [Fadigas e Pereira \(2013\)](#) a justaposição é o processo em que dois subgrafos

³corresponds to a complete subgraph [Barabási \(2016\)](#) p.324.

estão ligados por um único vértice e a sobreposição é quando os subgrafos estão conectados por dois ou mais vértices em comum.

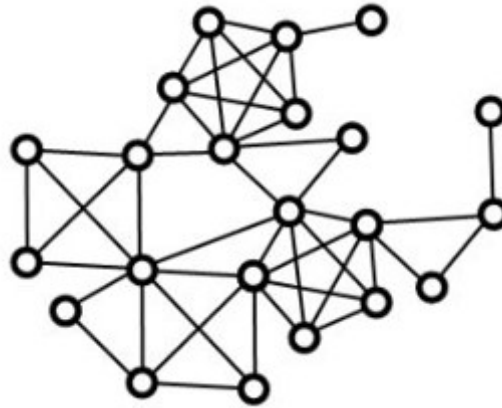


Figura 2.6: Formação das redes de cliques por justaposição e superposição. Fonte: [Fadigas e Pereira \(2013\)](#) p.2557.

Para elucidar ainda mais o conceito de cliques vejamos o exemplo com a seguinte frase: “Capoeira é bom. Os alunos amam capoeira”. Após os devidos tratamentos manuais e computacionais as frases ganham esses novos formatos: “Capoeira ser bom. Aluno amar capoeira” e quando passamos para visualização em redes as frases ficam representadas como na Figura 2.7

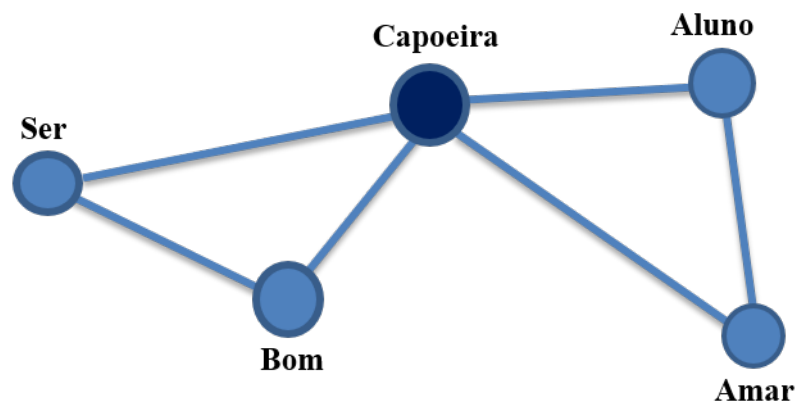


Figura 2.7: Formação das redes de cliques por justaposição. Fonte: Autora.

Em tempo, acreditamos que ao associarmos este conhecimento sobre RSBC com o aprofundamento dos estudos de comunidade e modularidade em redes apoiaremos o método da AC de maneira relevante.

2.4 Comunidades

Uma característica interessante passível de ser observada em diversas redes complexas é a existência de comunidades. Um fator relevante sobre as comunidades é que seus vértices compartilham características similares, possibilitando, assim, uma análise mais acurada dos dados, além de proporcionar comparações entre os grupos formados por conta da sua forte semelhança ou forte diferença. A Figura 2.8 representa o padrão de ligações de celular dos consumidores da maior empresa de telefonia móvel da Bélgica.

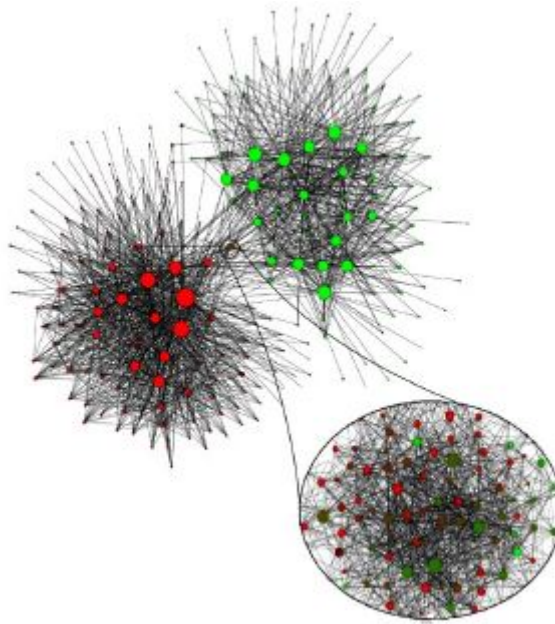


Figura 2.8: Rede de telefonia móvel belga. Fonte: B. Jean-L. G. e L. (2008, p.8).

Pelo fato dos vértices em comunidade possuírem fortes semelhanças ou diferenças consequentemente conseguimos perceber fenômenos de interações e prever tendências em processos nas questões subjetivas dos documentos analisados, sendo assim, Barabási (2016) define comunidade como “um grupo de nós que têm maior probabilidade de se conectar uns aos outros do que aos nós de outras comunidades”⁴.

Ainda segundo Barabási (2016) há duas definições básicas de comunidades: “comunidades fortes” e “comunidades fracas”, além de quatro Hipóteses sobre a formação e detecção das comunidades nas redes.

Isto posto, **comunidades fortes** são subgrafos (ou subgrupos) conectados onde seus vértices têm mais conexões com vértices na mesma comunidade que para vértices que

⁴In network science we call a community a group of nodes that have a higher likelihood of connecting to each other than to nodes from other communities.(BARABÁSI, 2016)

formam outras comunidades. Vejamos que o subgrafo C só é considerado uma comunidade forte quando os vértices dentro do subgrafo C tiver mais conexões dentro da comunidade do que com o resto da rede. Barabási (2016) diz, exatamente, que um subgrafo C é uma comunidade forte se para cada vértice $i \in C$ o grau interno do vértice i for maior que zero.⁵ A Equação 2.4:

$$k_i^{\text{int}}(C) > k_i^{\text{ext}}(C) \quad (2.4)$$

Ainda conforme Barabási (2016) as **comunidades fracas** são caracterizadas como subgrafos onde o grau interno total dos vértices ultrapassa o seu grau externo total da rede. Segundo Barabási (2016) o subgrafo C é uma comunidade fraca se o grau interno total de um subgrafo ultrapassar o seu grau externo total. Equação 2.5 comunidade fraca

$$\sum_{i \in C} k_i^{\text{int}}(C) > \sum_{i \in C} k_i^{\text{ext}}(C) \quad (2.5)$$

Ainda para apoiar e aprofundar os estudos sobre comunidades, devemos conhecer as Hipóteses que fundamentam o conceito e identificação das comunidades em redes.

- Hipótese 1 (H1) ou Hipótese Fundamental⁶: A primeira hipótese nos diz que a estrutura da comunidade na rede é representada através das redes e suas conexões. De acordo esta hipótese, existe uma verdade primeira sobre a organização das comunidades de uma rede, que pode ser observado pela estudo de sua matriz A_{ij} .
- Hipótese 2 (H2) ou Hipóteses de Conectividade e Densidade⁷: A H2 possui dois conceitos que a estruturam: o de “Conectividade” e “Densidade”, isso significa que uma comunidade é um conjunto de vértices conectados localmente densos na rede. Isto é, todos os vértices de uma comunidade devem ser alcançados através de outros vértices da mesma comunidade, pelas suas conexões (H2 - Conectividade). Concomitantemente, esperamos que os vértices da mesma comunidade tenham maiores chances de se ligar aos outros vértices da sua comunidade do que aos vértices de outras comunidades (H2 - Densidade).
- Hipótese 3 (H3) ou Hipótese Aleatória⁸: Em uma rede aleatória, o padrão de ligações entre os vértices devem ser uniforme, independente da distribuição de grau da rede.

⁵subgraph C forms a strong community if for each node $i \in C$. (BARABÁSI, 2016, p.326).

⁶H1: Fundamental Hypothesis. (BARABÁSI, 2016).

⁷H2: Connectedness and Density Hypothesis. (BARABÁSI, 2016).

⁸H3: Random Hypothesis. (BARABÁSI, 2016)

Em conseqüente, as redes aleatórias carecerem de alterações de densidade locais sistemáticas, ou seja, de uma estrutura de comunidade inerente. Essas questões supracitadas inspiraram a terceira hipótese, que compara “a densidade de ligação de uma comunidade com a densidade de conexões obtidas para o mesmo grupo de vértices para uma rede aleatoriamente redesenhada, assim poderíamos decidir se a comunidade original corresponde a um subgrafo denso ou seu padrão de conectividade surgiu por acaso”. (BARABÁSI, 2016)⁹.

- Hipótese 4 (H4) ou Hipótese de Modularidade Máxima¹⁰: Para um determinada rede a partição com máxima modularidade corresponde à estrutura da melhor comunidade. Esta hipótese é o ponto de saída de vários algoritmos de detecção de comunidades, cada algoritmo visa buscar partição com maior modularidade (M).

É possível identificar a melhor partição ao verificar a M para todas as partições possíveis, escolhendo aquela para qual M foi o maior. Mas o número muito grande de partições, através dessa abordagem computacionalmente é inviável (BARABÁSI, 2016)¹¹. Diversos autores na última década buscaram elaborar algoritmos que encontram partições com a modularidade próxima ao máximo, sem a necessidade de inspecionar todas as partições existentes na rede.

Na próxima Subseção 2.4.1 abordaremos os conceitos e finalidades da modularidade e os principais algoritmos de identificação de comunidades nas redes.

2.4.1 Modularidade

Nas redes complexas, não raro é possível identificar grupos ou comunidades, que também são chamados de *clusters*. Inclusive nas redes aleatórias, com o auxílio da H3, é possível identificar comunidades. Há diversas formas para identificar comunidades em redes, estratégias divisivas de partições e o índice que avalia a qualidade destas partições é o da **modularidade** (M).

Segundo Barabási (2016), uma rede com N vértices e L arestas e uma partição em comunidades n_c , cada comunidade com vértices N_c conectados entre si por Arestas L_c , onde

⁹This hypothesis has some actionable consequences: By comparing the link density of a community with the link density obtained for the same group of nodes for a randomly rewired network, we could decide if the original community corresponds to a dense subgraph, or its connectivity pattern emerged by chance.(BARABÁSI, 2016)

¹⁰H4: Maximal Modularity Hypothesis. (BARABÁSI, 2016)

¹¹The maximum modularity hypothesis is the starting point of several community detection algorithms, each seeking the partition with the largest modularity. In principle we could identify the best partition by checking M for all possible partitions, selecting the one for which M is largest. Given, however, the exceptionally large number of partitions, this brute-force approach is computationally not feasible. Next we discuss an algorithm that finds partitions with close to maximal M , while bypassing the need to inspect all partitions.(BARABÁSI, 2016)

$c = 1, \dots, n_c$. Se L_c for maior que o número esperado de conexões entre os vértices N_c dada a sequência de grau da rede, então os vértices do subgrafo C_c podem ser parte de uma comunidade verdadeira, conforme esperado com base na Hipótese de Densidade H2. Portanto, medimos a diferença entre as conexões reais da rede (A_{ij}) e o número esperado de conexões entre i e j se a rede for conectada aleatoriamente (p_{ij}), matematicamente expressa na Equação 2.6:

$$M_c = \frac{1}{2L} \sum_{(i,j) \in C_c} (A_{ij} - p_{ij}) \quad (2.6)$$

Aqui, p_{ij} pode ser determinado ao aleatorizar a rede original, mantendo o grau esperado de cada vértice inalterado. Usando o modelo nulo de grau preservado, temos a Equação 2.7:

$$p_{ij} = \frac{k_i k_j}{2L} \quad (2.7)$$

Ainda segundo Barabási (2016) se M_c for positiva, então o subgrafo C_c possui mais arestas do que o esperado aleatoriamente, portanto, representa uma comunidade em potencial. Se M_c for zero, então a conectividade entre os vértices N_c é aleatória, explicada pela distribuição de grau. Finalmente, se M_c for negativo, então os vértices de C_c não formam uma comunidade.

Percebamos que quanto maior for o valor de M melhor será aquela partição. Isto posto, podemos definir a modularidade em quatro conceitos para como ilustrado na Figura 2.9:

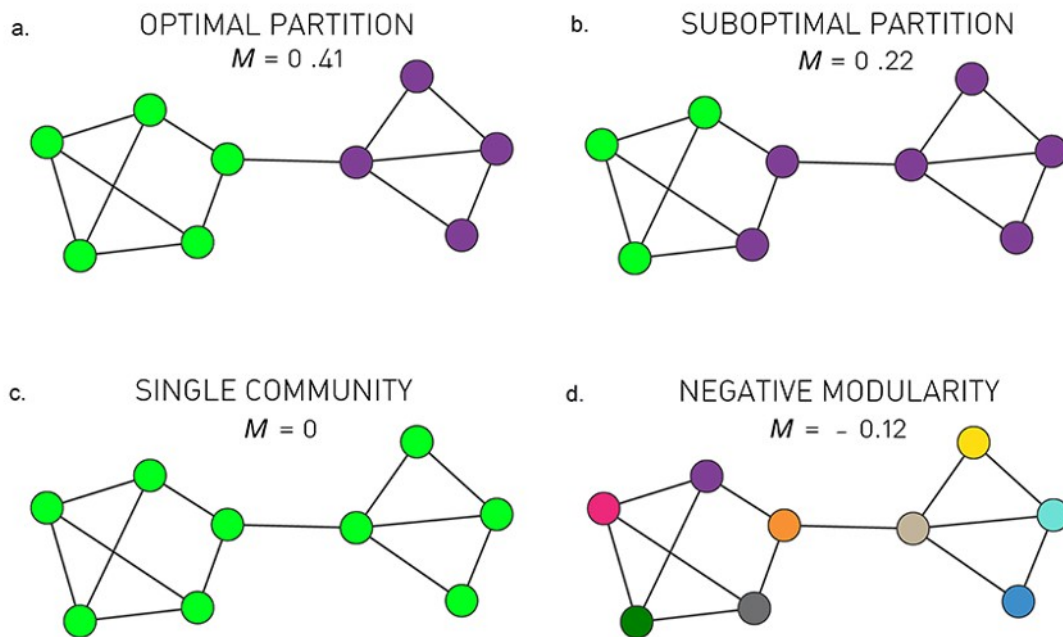


Figura 2.9: Definições modularidades. Fonte: [Barabási \(2016, p.340\)](#).

- Partição Ótima: A partição com modularidade máxima onde $M = 0,41$.
- Partição Subótima: Uma partição com uma modularidade não tão perfeita, mas ainda sim positiva, $M = 0,22$.
- Comunidade única: Se atribuímos todos os vértices à mesma comunidade, obtemos $M = 0$, independentemente da formação da rede.
- Modularidade negativa: Se atribuímos cada vértice a uma comunidade diferente, a modularidade é negativa, obtendo $M = -0,12$.

Para um conhecer mais detalhes sobre modularidade recomendamos a leitura do livro [Barabási \(2016\)](#). Na próxima Seção falaremos brevemente sobre os principais tipos de algoritmos para identificação de comunidades em redes.

2.4.2 Algoritmo de Louvain

A otimização da modularidade oferece uma abordagem inovadora para a detecção da comunidade em redes, pois muitos algoritmos com boa precisão nas partições das comunidades, geram em contrapartida um alto custo computacional, muitas vezes tornando inviável seu uso em redes grandes.

O *software* (*Gephi* 0.9.1) que utilizamos para o cálculo da modularidade tem implementado em seu sistema o algoritmo “Louvain” alaborado pelos pesquisadoress [B. Jean-L. G. e L. \(2008\)](#) e seus colaboradores. Este algortimo é capaz de trabalhar com redes grandes (i.e vértices na casa dos milhares) com um custo computacional baixo. Basicamente o algoritmo de Louvain é dividido em duas etapas.

Etapa I: A modularidade é otimizada por mudanças locais. É escolhido um vértice e calculado a mudança na modularidade, se o vértice se juntar à comunidade de seus vizinhos imediatos. A figura mostra a mudança de modular esperada $\Delta M_{o,i}$ para o vértice 0. Assim, o vértice 0 unirá o vértice 3, pois a mudança de modularidade para este movimento é a maior, sendo $\Delta M_{0,3} = 0,032$. Este processo é repetido para cada vértice, concluindo a Etapa I ([BARABÁSI, 2016, p.371](#)).

Etapa II: As comunidades obtidas no Etapa I são agregadas, criando uma nova rede de comunidades. Os vértices pertencentes à mesma comunidade são mesclados em um único vértice, como mostrado no canto superior direito na Figura [2.10](#). Esse processo gerará autoloops, correspondendo as conexões entre vértices na mesma comunidade que agora são incorporados em um único vértice ([BARABÁSI, 2016, p.371](#)). A Equação [2.8](#) representa as duas Etapas algoritmo Louvain:

$$\Delta M = \left[\frac{\Sigma_{in} + 2k_{i,in}}{2W} - \left(\frac{\Sigma_{tot} + k_i}{2W} \right)^2 \right] - \left[\frac{\Sigma_{in}}{2W} - \left(\frac{\Sigma_{tot}}{2W} \right)^2 - \left(\frac{k_i}{2W} \right)^2 \right] \quad (2.8)$$

A soma das Etapas I e II é chamada de passagem. A rede obtida após cada passagem é processada novamente (Passagem 2), até que não seja possível aumentar a modularidade. Como ilustrado na Figura [2.10](#)

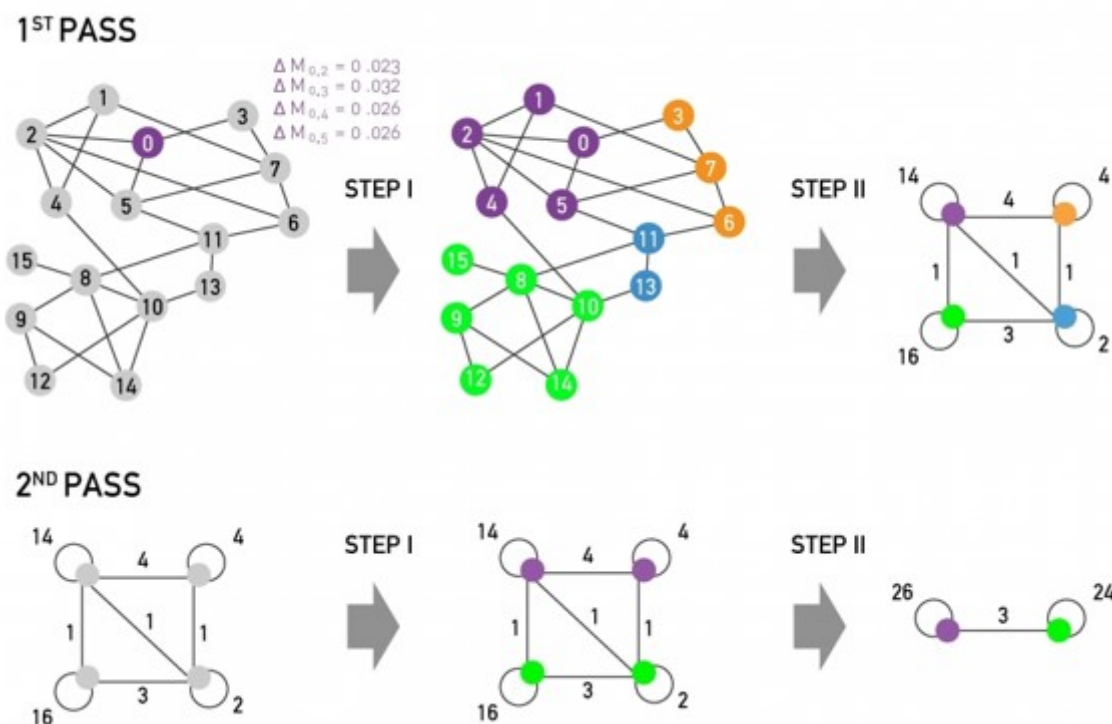


Figura 2.10: Etapas do procedimento do algoritmo de Louvain. Fonte: Barabási (2016, p.371).

2.5 Análise de Conteúdo

Laurence Bardin é professora assistente de Psicologia na Universidade de Paris V, utilizou as técnicas da Análise de Conteúdo na investigação psicossociológica e no estudo das comunicações de massas. Segundo a autora, a análise de conteúdo pode ser conceituada como um conjunto de técnicas para a análise das comunicações visando alcançar no conteúdo do discurso indicadores que possibilitem inferências de conhecimentos relativos às condições de produção e recepção dessas mensagens.

O analista labora com palavras que, isoladas, podem conferir relações com a mensagem ou viabilizar a inferência de conhecimento a partir da mensagem. São estabelecidas relações entre as estruturas semânticas e as estruturas psicológicas ou sociológicas dos enunciados. A percepção do pesquisador responsável pela análise não é, por conseguinte, uma leitura à letra, mas, o enaltecimento de um sentido que está em segundo plano (BARDIN, 1977). A análise categorial é um dos procedimentos utilizados para a realização da análise de conteúdo, e, em seu desenvolvimento, secciona o discurso em categorias.

2.5.1 Principais características da Análise de Conteúdo

A AC é um método de tratamento e análise de informações, extraídas por meio de técnicas de coleta de dados, aplicada em discursos ou qualquer comunicação reduzida a um texto ou documento. Tem como finalidade a compreensão do sentido das comunicações, o discurso dos indivíduos, do seu conteúdo objetivo ou subjetivo, tendo como base a decodificação dos dados por meio da decomposição do texto em unidade léxica ou categorial.

Esse procedimento é caracterizado como um método de tratamento da informação contida nas mensagens. Para sua execução é imprescindível a sistematização das unidades de registro (UR) e categorias (CAT) relacionados ao objeto estudado. As inferências que serão geradas a partir das categorias são responsáveis pela identificação das questões relevantes contidas no conteúdo das mensagens.

Para [Bardin \(2010\)](#), a realização da análise de conteúdo, requer a identificação dos temas tratados nos documentos para definir as unidades de registro, e posteriormente as categorias.

É oportuno citar que para amparar o caminho entre o agrupar das UR em CAT, os pesquisadores [Cunha \(2012\)](#) e Núbia Moura Ribeiro, elaboraram um procedimento que organiza as UR em subgrupos que possuem afinidades conceituais, morfológicas ou sintáticas fazendo emergir assim os Núcleos de Sentidos (NS). Os NS geram um refinamento nos dados e auxiliam nas inferências do analista por especificar ainda mais a CAT que eles compõem. Vale lembrar que por mais eficientes que sejam os NS, eles não são originários da proposta metodológica de [Bardin \(2010\)](#).

Assim sendo, [Bardin \(2010\)](#), nos diz que a AC possui três polos cronológicos retratados a seguir:

a) Pré-Análise: Caracterizada por ser uma fase intuitiva; da escolha dos documentos a serem submetidos à análise.; formulação de hipóteses e dos objetivos; a elaboração de indicadores que fundamentem a interpretação final, tratamento dos dados antes da análise.

- Ação Leitura “flutuante”: essa primeira atividade consiste em estabelecer contato com os documentos a analisar e em conhecer o texto. Tem a intenção de aos poucos afunilar a leitura, a deixando mais precisa, devido as suas hipóteses, objetivos e a natureza do material.
- Escolha dos documentos: a maneira que escolhemos os documentos para serem analisados. Há quatro regras que nos orientam: (1) Regra da exaustividade (não se

pode deixar de fora nenhum elemento que não se justifique o motivo); (2) Regra da representatividade: analisa a relevância da amostra; (3) Regra da homogeneidade: os documentos escolhidos devem obedecer critérios rigorosos de escolha; (4) Regra da pertinência: os documentos selecionados necessitam corresponder aos objetivos da pesquisa.

b) Exploração do material: Caracterizada por ser uma fase longa e fastidiosa, consiste na codificação, decomposição ou enumeração do documento analisado.

- Ação: (1) Elaboração das Unidades de Registro (UR); (2) Categorias (CAT).

c) Tratamento dos resultados, a inferência e a interpretação: Caracterizada por ser uma fase de inferências e interpretação dos dados.

- Ação: (1) Tratamento dos resultados brutos na intenção de deixá-los significativos e válidos. (2) A aplicação de operações e provas estatísticas (e.g. percentagens, análise fatorial) para serem inferidos e interpretados pelo analista.

Após realizar as três primeiras etapas acima citadas da AC, há uma necessidade de estudar com mais precisão o discurso (i.e. textos técnicos, literários, entrevistas transcritas) para destrinchá-lo e assim podermos visualizar e analisar melhor o que se procura identificar ou descobrir.

Dentre as etapas existentes na AC, a subdivisão das Unidades de Registro (adivindo das Unidades de Contexto) e a Categorização são relevantes. Seus conceitos estão explicados a seguir:

Unidade de Registro (UR)¹²: é uma unidade de significação a ser codificada e corresponde ao menor segmento do conteúdo a ser considerado como unidade de base, buscando à categorização, podendo ser de natureza e dimensões distintas (BARDIN, 2010).

Unidades de contexto (UC): são os segmentos de mensagens cujas dimensões permitem ao pesquisador compreender o significado da UR ou seja, o que o pesquisador consegue extrair das falas dos sujeitos. Contribui para a compreensão de sentidos a fim de codificar as unidades de registro que, agrupando-as, atribui-lhes um sentido engajado, ou seja, corresponde ao segmento da mensagem que, pela dimensão superior, propicia entender o significado de registro (BARDIN, 1977).

¹² “É necessário fazer a distinção entre a UR - o que se conta - e a regra de enumeração - o modo de contagem”. Bardin (2010). Existem seis regras de enumeração, a mais usual é a *frequência*, essa regra indica que a importância da palavra aumenta de acordo com a frequência de aparição. Para saber mais ver Bardin (2010).

Categorias (CAT): são rubricas ou classes de um conjunto de unidades de registro, sob um título genérico, em razão de características comuns [Andrade \(2013\)](#). Em síntese, a Análise de Conteúdo objetiva a compreensão do sentido da comunicação, por meio da decomposição do texto em unidade léxica ou categorial.

Modelo Computacional Proposto: redes semânticas de cliques

Nessa parte do trabalho detalharemos todas as etapas que abrangem a elaboração do modelo computacional da pesquisa, a amostragem e seus critérios, os métodos e a visualização final das redes geradas. Este capítulo traz três Seções, abordando os seguintes temas:

1. Método da Pesquisa;
2. Distribuição espacial das redes semânticas de cliques em comunidades;
3. O cenário das fontes dos dados.

3.1 Método da Pesquisa: Modelo Computacional Proposto

Os procedimentos metodológicos da nossa pesquisa estão subdivididos em cinco etapas ilustradas na Figura [3.1](#).

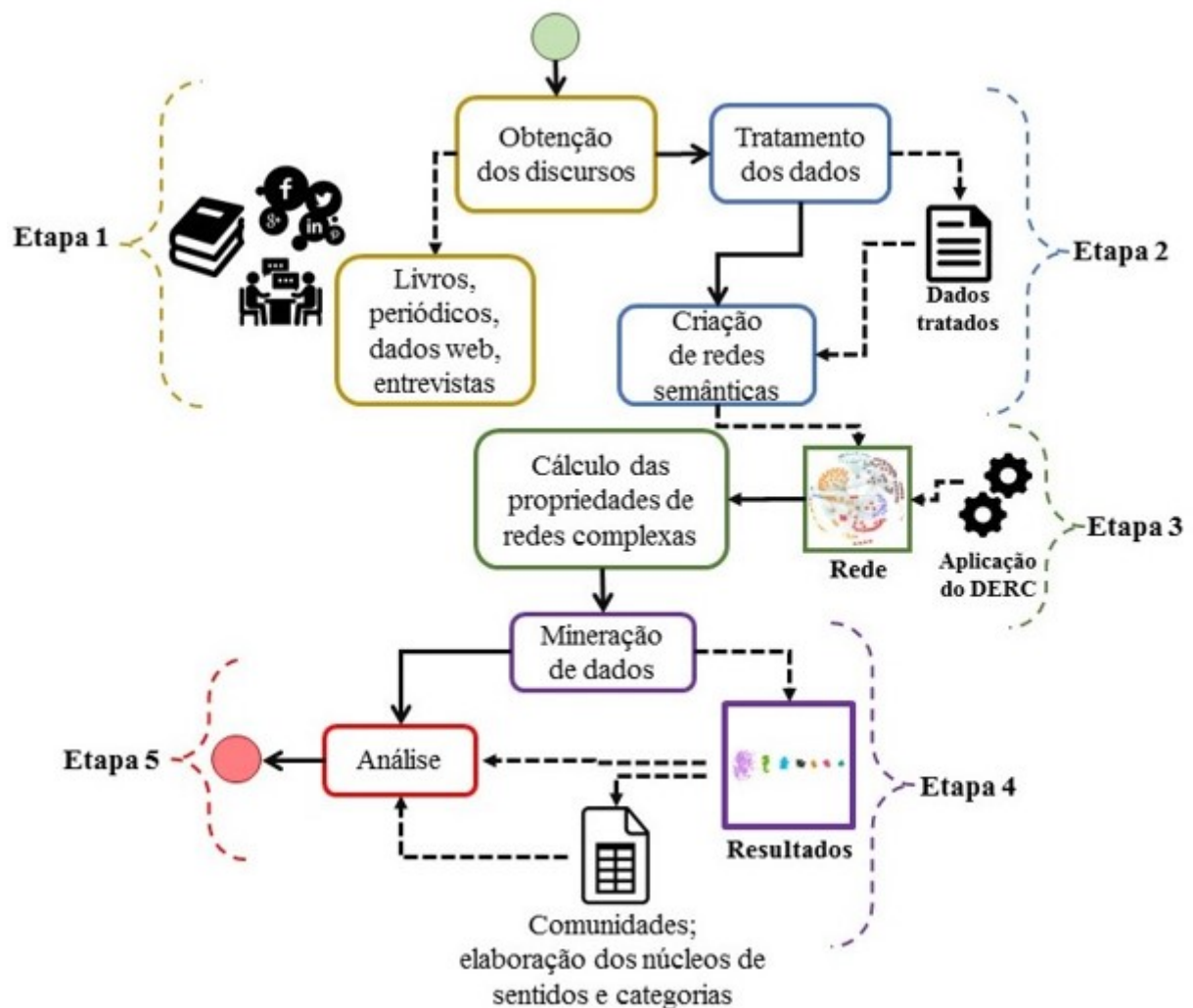


Figura 3.1: Fluxograma da metodologia Modelo computacional. Fonte: Autora.

Na **Etapa 1** (cor ocre da Figura 3.1) da metodologia do trabalho envolve: (1) a obtenção dos discursos do tema escolhido para estudo, esses que possuem origens dos mais diversos ambientes: artigos periódicos, livros, relatórios técnicos, entrevistas. Vale lembrar que a obtenção ou acesso ao discurso é fundamental para o prosseguimento do trabalho.

Na **Etapa 2** (cor azul da Figura 3.1) retrata o método de construção de redes semânticas (RS) que consiste basicamente na (1) eliminação das palavras sem significados relevantes e conectivos (2) e na alteração das palavras restantes para sua forma canônica, como seguem Caldeira (2005) e Teixeira (2007). Após a realização da primeira etapa supracitada, (3) utilizamos o método sobre RS baseadas em cliques, dos estudos de Fadigas e Pereira (2013).

Em seguida, na **Etapa 3** (cor verde da Figura 3.1), devemos utilizar o algoritmo de

modularidade¹ implementado no *software Gephi 0.9.1*² para a detecção das comunidades, além dos cálculos dos demais índices de redes (i.e. densidade, coeficiente de aglomeração, caminho mínimo médio). Para finalizar a Etapa 3 aplica-se a Distribuição espacial das redes semânticas de cliques em comunidades (DERC), um procedimento que proporciona melhor visualização dos vértices ou comunidades desejadas.

A **Etapa 4** (cor roxa da Figura 3.1) representa os resultados parciais gerados através da mineração de dados que permite a perfeita identificação das Comunidades, além da elaboração dos Núcleos de Sentido e Categorias.

Por fim, a **Etapa 5** (cor vermelha da Figura 3.1) consiste na análise dos resultados de forma mais refinada (conclusivas) e inferências do analista.

Vale ressaltar que o nosso trabalho não pretende discutir as questões de análise subjetiva, intrínsecas nos dois cenários de dados escolhidos: AC realizada em uma tese doutoral e relatório de educação da UNESCO.

3.1.1 *Distribuição espacial das redes semânticas de cliques em comunidades (DERC)*

As comunidades em redes são imprescindíveis para o estudo de grupos. No entanto, uma desvantagem para análise é quando há muitos vértices na rede, a visualização espacial da rede fica prejudicada, conseqüentemente compromete a visualização das comunidades e posterior análise.

Vale lembrar, que utilizamos as RS de cliques por acreditarmos que para esse estudo utilizar a sentença como a menor unidade de um texto, facilita a formação das comunidades e conseqüentemente dos NS e CAT. Para tanto, desenvolvemos um procedimento de distribuição e visualização espacial da rede no *software Gephi 0.9.1* (i.e. *software* escolhido como ferramenta de visualização de redes para a nossa pesquisa).

A Figura 3.2 descreve as etapas necessárias para realizar o DERC.

¹O algoritmo de modularidade usado pelo *Gephi* são dos autores B. Jean-L. G. e L. (2008).

²Programa de visualização e cálculos dos índices de redes

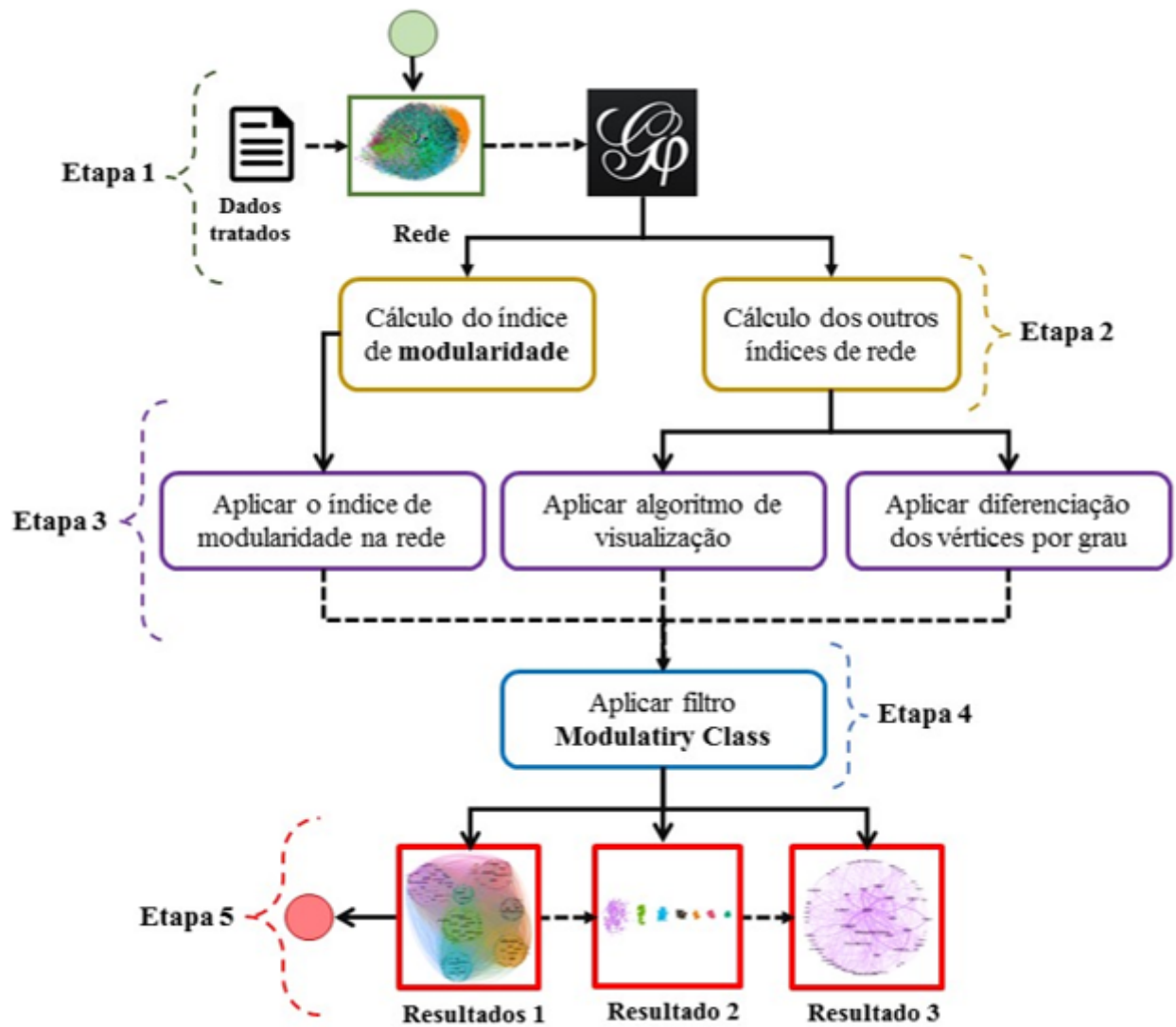


Figura 3.2: Fluxograma da metodologia do DERC. Fonte: Autora.

Na **Etapa 1** (cor verde da Figura 3.2), após carregar o arquivo *.NET* no *Gephi*, é necessário calcular a modularidade da rede, retirar os pesos das arestas³, permanecer com a resolução em 1.0⁴ e selecionar a opção “aleatória”.

Na **Etapa 2** (cor ocre da Figura 3.2) para uma visualização mais refinada das comunidades geradas, pela modularidade do *Gephi*, recomenda-se ir na aba “aparência”, na opção “nós”, e em “partição” selecionar o atributo “modularidade” e aplica-lo. Deste modo as comunidades aparecerão pintadas na rede. Vale ressaltar que a qualquer momento podemos alterar as cores das comunidades.

³para o nosso estudo o peso das arestas não é relevante

⁴após alguns teste verificamos que a resolução em 1.0 melhor subdivide a rede em comunidades

Nesta **Etapa 3** (cor roxa da Figura 3.2) é necessário ir no ícone de “tamanho”, selecionar o Grau para diferenciar o tamanho dos vértices e aplicar. Em seguida, apesar de não haver no *Gephi* um algoritmo específico para separar a rede por comunidades, é válido utilizar um algoritmo como o *Atlasforce*, porque segundo [Paixao \(2016\)](#) esse é um modelo de algoritmo de vetor de força que segue o princípio da força atrativa e de repulsiva para determinar o posicionamento dos vértices no espaço.

Os vértices não possuem locais fixos, eles são posicionados relativamente aos outros vértices atendendo a fórmula do algoritmo de atração e repulsão. Essas três primeiras Etapas são as mais utilizadas para distribuição espacial dos vértices na rede e mesmo assim não favorecem de maneira relevante a visualização da rede como exemplificado na Figura 3.3.



Figura 3.3: Exemplo das 3 primeiras Etapas da metodologia do DERC. Fonte: Autora.

A seguir continuaremos a demonstrar onde o DERC se sobressai.

A **Etapa 4** (cor azul da Figura 3.2) é crucial para diferenciar as comunidades e assim estudá-las individualmente, pois mesmo passando pelos procedimentos supracitados, ainda não conseguimos de maneira satisfatória definir as comunidades.

Por conta do impasse de distinguir as comunidades é necessário aplicar um filtro. Devemos nos dirigir para a aba de “Filtros”, “Biblioteca”, “Atributos”, “números de partições” e arrastar a opção *Modulatory Class* para o local indicado. Logo embaixo um *slide* de cor azul aparecerá. Na medida em que o deslizamos as comunidades são incluídas ou retiradas.

Na Etapa 5 (cor vermelha da Figura 3.2) a Figura 3.4 representa o que desejamos alcançar ao fim da aplicação do DERC. O DERC possibilita o estudo individual de cada comunidade de maneira clara, pois a partir dele, por exemplo, é possível calcular os índices de rede de cada comunidade, e não somente de toda a rede. E se for necessário, é possível executar todo a DERC em somente um grupo de vértices, ou seja, detectar subcomunidades na comunidade.

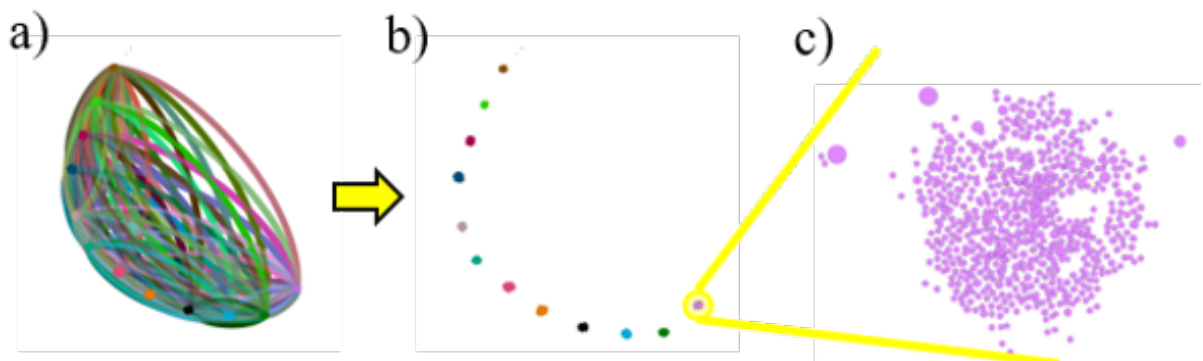


Figura 3.4: Essa imagem representa todo o processo da DERC. Fonte: Autora.

Por fim, acreditamos que esse processo de Distribuição espacial das redes semânticas de cliques em comunidades apoia a AC no que diz respeito a elaboração dos NS e CAT. Trataremos desse assunto de forma mais aprofundada no Capítulo 4.

3.2 Fonte de dados, os seus cenários e os instrumentos utilizados

Nossa primeira fonte de dados refere-se a uma pesquisa doutoral [Andrade \(2013\)](#) que trabalhou com AC e RS. [Andrade \(2013\)](#) coletou os discursos de pesquisadores e/ou docentes da área de “Colaboração científica e Difusão do Conhecimento” através de uma entrevista semiestruturada. Já a nossa segunda fonte é um relatório técnico de educação (Educação para Todos - EPT) da UNESCO do ano de 2015.

Vale lembrar que o nosso principal objetivo na pesquisa é a elaboração de um modelo computacional que permita apoiar de forma relevante o processo da AC. Sendo assim, com a intenção de minimizar e/ou tendenciar os nossos resultados, optamos em utilizar um banco de dados onde houve a realização da AC e já teria sido validado pela comunidade científica. A segunda aplicação do nosso modelo tem como finalidade oferecer outro meio para a elaboração dos NS (proposto por [Cunha \(2012\)](#) e utilizado por [Andrade \(2013\)](#)) e CAT da AC a partir das RSC .

3.2.1 Cenário 1: tese doutoral

A fonte de dados da autora [Andrade \(2013\)](#) foram docentes e/ou pesquisadores de programas de pós graduação e o instrumento escolhido foi uma entrevista semiestruturada, com perguntas relacionadas a área de “Colaboração científica e Difusão do Conhecimento relativas”. Ao todo foram realizados 17 entrevistas e foi aplicada a AC. Com a execução da AC foram geradas 3 Categorias e 7 Núcleos de Sentido mostrados na Tabela [3.1](#)

Tabela 3.1: Categorias e Núcleos de Sentido do cenário 1. (Fonte:[Andrade \(2013\)](#) p.76)

Categorias	Núcleo de Sentido/significado
Difusão do Conhecimento e Colaboração	Difusão do Conhecimento
	Colaboração
	Motivação e Condutas
Estruturas Acadêmicas	Academia ou Universidade
	Comunidade Acadêmica
Políticas e CTI	Políticas de Ciência e Tecnologia
	Políticas de Publicação

O critério inicialmente de categorização adotado foi o morfológico (i.e. análise da formação das palavras) e o semântico, com base nas categorias e seus temas associadas aos objetivos da pesquisa [Andrade \(2013\)](#).

É válido frisar que na segunda fase na AC (i.e exploração do material), se faz o recorte do material analisado, esse ação possibilita a definição de unidades de registro e de contexto. Sendo assim, seguiremos o recorte realizado pela professora [Andrade \(2013\)](#), como ilustrado a Tabela ??

Tabela 3.2: 10 palavras com maior frequência do NS1 e NS2. [Andrade \(2013, p.100 e p.113\)](#).

Unidade de registro	Frequência	Unidade de registro	Frequência
Publicação	185	Parceria	71
Pesquisa	172	Colaboração	68
Conhecimento	126	Vínculo	64
Área	98	Coautoria	60
Artigo	69	Banca	59
Projeto	68	Relação	48
Empecilho	65	Rede	44
Difusão	62	Benefício	40
Difusão do conhecimento	46	Compartilhar	36
Trabalho	46	Afinidade	33

É importante ressaltar que a professora [Andrade \(2013\)](#), ao elaborar os NS, concentrou-se em estudar somente as 10 primeiras Unidades de Registro (UR) de cada um dos sete Núcleos de Sentido. Assim sendo, os nossos esforços de análise dos vértices e aplicação do nosso Modelo Computacional na rede semântica seguirá o rigor do recorte do banco de dados da pesquisa doutoral da professora [Andrade \(2013\)](#).

3.2.2 Cenário 2: Relatório de Monitoramento Global: EPT 2015

A escolha da segunda fonte de dados advém da importância que a educação tem perante as demandas sociais atuais, significado que a UNESCO diz ter para o mundo. Na virada do século XX para o XXI, no Fórum Mundial de Educação em Dakar, Senegal, 164 chefes de estados concordaram com o “Marco de Ação de Dakar”, ou seja, “lançou uma agenda ambiciosa para alcançar seis objetivos educacionais de amplo alcance até 2015” ([UNESCO, 2015](#)).

Em apoio, a UNESCO iniciou os Relatórios de Monitoramento de EPT (Educação Para Todos) no ano de 2002, para monitorar o progresso da educação, alertar lacunas persistentes e fazer recomendações para a agenda global de desenvolvimento sustentável posterior a 2015. Na Figura 3.5, podemos ver todos os EPTs lançados de 2002 - 2015.



Figura 3.5: Capas dos EPT já realizados. Fonte: retirado do site da UNESCO.org.

O relatório escolhido como banco de dados foi “Educação Para Todos - 2000-2015: progressos e desafios”, pela fato do EPT 2015 englobar os 15 anos de monitoramento. Para o tratamento dos dados e elaboração das redes semânticas aplicamos os procedimentos já descritos na Seção 2.3.

No próximo Capítulo 4, abordaremos a convergência entre a AC e as RS, as inferências, o processo de DERC e aplicação do modelo computacional proposto na elaboração e acurácia dos Núcleos de Sentido e Categorias dos dois cenários de dados.

Resultados e Discussões

Neste capítulo, retrataremos detalhadamente todas as etapas que dizem respeito a discussão, execução e explanação do modelo computacional proposto, até os cálculos dos parâmetros analisados da AC e RS.

4.1 *Propriedades convergentes na Análise de Conteúdo e das Redes Semântica de Cliques*

Por detrás de todo o procedimento do Capítulo 3 há um estudo mais específico das propriedades em redes, e para a nossa dissertação voltamos a atenção para os índices que permitem o apoio ao método da análise de conteúdo de Bardin, especificamente para a elaboração dos NS e CAT.

A convergência (apoio) das RS para a AC consiste em percebermos as similaridades dos seus elementos e as propriedades que os caracterizam, e sugerir novos procedimentos para identificação e formação das UC, NS e CAT da AC.

a) Primeira similaridade - A análise de conteúdo pode ter como menor unidade de estudo a “palavra” (Unidade de Registro - UR); geralmente nas RS também trabalha com a “palavra” (vértice) sendo a menor unidade da rede.

Neste caso de estudo as UR são classificadas pela sua frequência de aparição no discurso, e no caso das RS algo similar acontece, quando descobrimos o *Grau* do vértice, que representa o número de conexões que ele tem com os demais elementos do discurso.

No entanto, o cálculo do *Grau* é feito de maneira computacional, no caso da frequência da AC, não raro, é realizado de forma manual levando, desde o início, um trabalho árduo.

Cabe ressaltar que nas duas teorias, somente conhecer as UR ou os vértices não nos dizem muito, precisamos compreender o contexto em que esses elementos estão inseridos.

b) Segunda similaridade - Para conhecermos o contexto da UR, utilizamos a Unidade de Contexto (sentença) para podermos compreender o cenário que a UR se encontra. Teremos que repetir esse processo de análise e procura da UC **todas às vezes** em que a UR for contabilizada.

Nas redes semânticas baseadas em cliques, uma palavra (vértice) que aparece em uma *Clique*, já está inserida no seu próprio contexto. Pois, as *Cliques* são formadas por sentenças.

Não há necessidade de procurar no discurso “n” vezes a UC, pois o contexto já estaria ilustrado nos vizinhos do vértice analisado. Facilitando de maneira significativa essa parte da análise, pois basta analisar a *Clique* para conhecer sentença (ou contexto) em que a Unidade de registro está inserida.

- c) **Terceira similaridade** - Na AC após registrar as UR, essas são agrupadas em áreas temáticas relacionadas às Unidades de Contexto, originando os Núcleos de Sentido relacionados às categorias. Novamente outro procedimento manual, onde o trabalho aumenta proporcionalmente a extensão do documento analisado.

Nas RS baseadas em *Cliques*, as Comunidades geradas pelo algoritmo de Modularidade sugestionam as Categorias e o conjunto de *Cliques* que as compõem, analogamente sugerem os Núcleos de Sentido.

Essa afirmação é possível pelo fato de que as comunidades são organizadas por “afinidades”, ou seja, por propriedades matemáticas (i.e. densidade, diâmetro, grau, modularidade) que sugerem a melhor distribuição dos grupos. Vejamos a Figura 4.1:

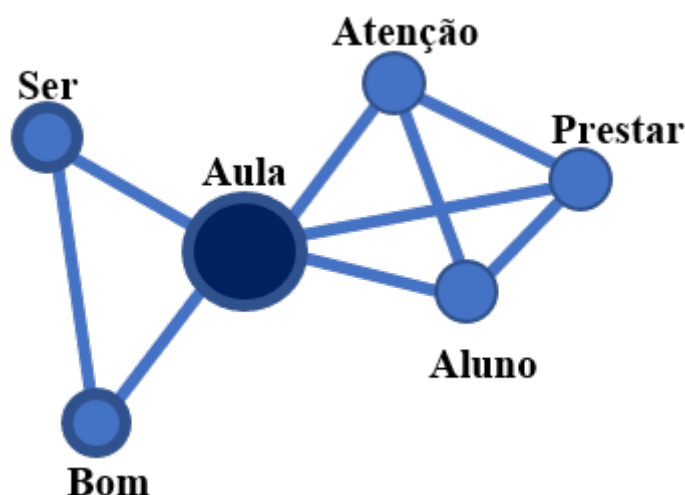


Figura 4.1: Uma rede semântica composta por duas cliques por justaposição. Fonte: Autora.

- d) **Exemplo** - Na Figura 4.1, imaginemos que a UR analisada seja “Aula” com frequência igual a 2 e grau $\langle k \rangle = 5$, ou seja, não há palavra com maior frequência que “Aula”, nem com maior grau.

Ao visualizar este discurso através das RSBC, podemos perceber de maneira quase que imediata as sentenças (Unidades de Contexto) que a palavra “Aula” pertence no discurso. O recurso das redes nos aponta indícios das UC, assim só precisaríamos recorrer ao texto em caso de dúvidas ou para ratificar a sentença que a palavra se

encontra. Nesse simples exemplo podemos perceber desde já um apoio importante do uso de redes na AC.

4.2 Aplicação do Modelo Computacional Proposto: Análise experimental/Simulações

Com o objetivo de averiguar o nosso modelo, fizemos uma simulação em duas amostras: (1) um discurso já analisado pela análise de conteúdo; (2) um relatório técnico sobre educação com 58 páginas. Aplicamos o nosso modelo computacional nos cenários afim de verificar se o modelo proposto seria preciso em indicar NS e CAT para a Análise de Conteúdo.

4.2.1 Simulação no Cenário 1

Só foi possível realizar a primeira simulação pelo fato de termos tido acesso a planilha de dados das UR, NS e CAT dos discursos da tese doutoral.

A seguir a Figura 4.2 ilustra os procedimentos que utilizamos para realizar a simulação e posterior análise referente ao apoio das redes semânticas de cliques à análise de conteúdo.

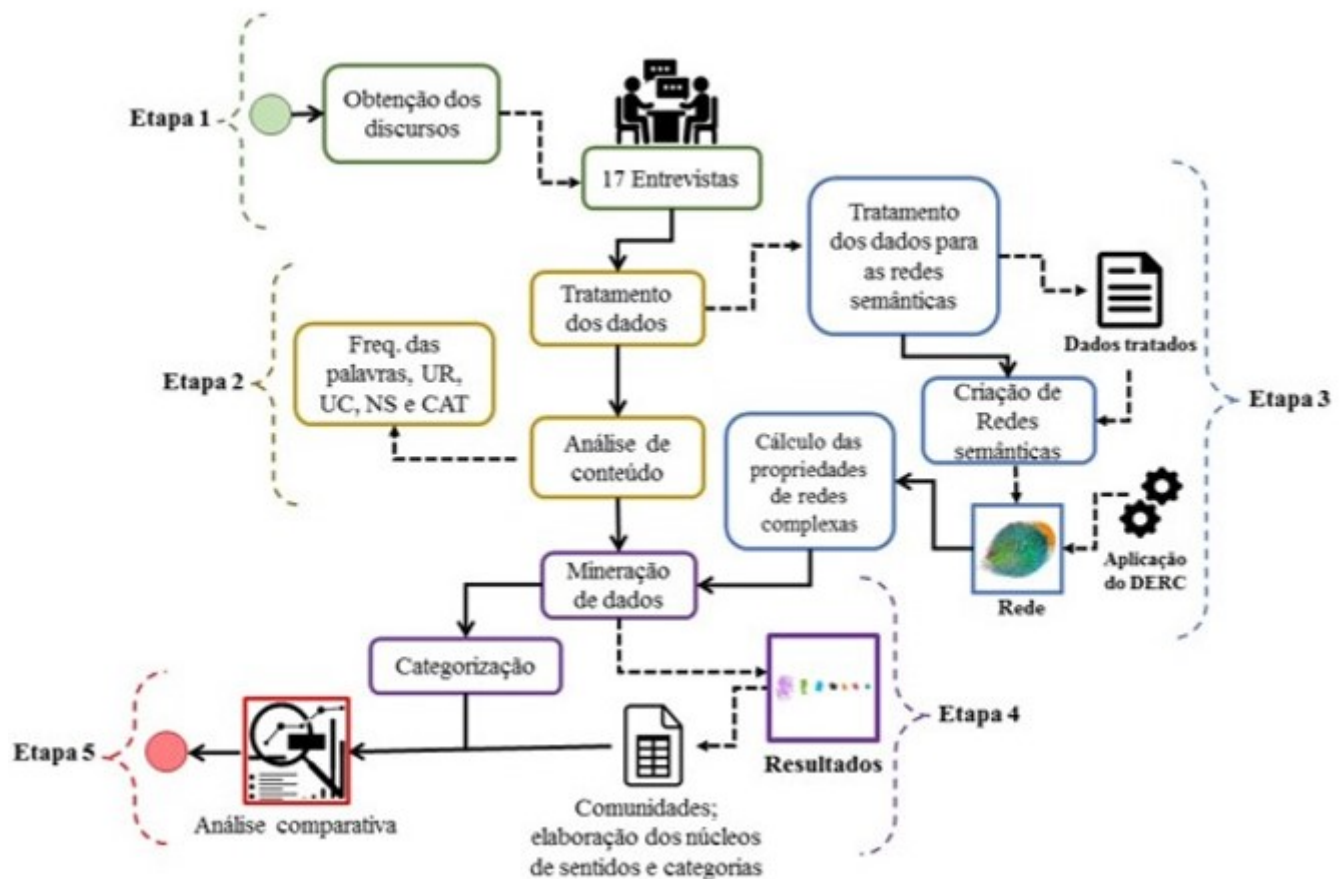


Figura 4.2: Fluxograma da aplicação do MC proposto no primeiro cenário da amostragem. Fonte: Autora.

A Etapa 1 (cor verde da Figura 4.2) foi totalmente realizada pela autora Andrade (2013) onde os sujeitos da pesquisa são docentes de Programa de Pós Graduação, ao todo foram 17 entrevistas, onde 12 pessoalmente e 5 via o *Skype*¹.

Na Etapa 2 (cor ocre da Figura 4.2) que consiste na sistematização, tratamento dos dados e aplicação da AC “foi concretizada com a ajuda de planilhas, nas quais foram registradas as unidades de registro extraídas das entrevistas classificadas e, a partir daí, estas unidades de registro foram agrupadas em áreas temáticas relacionadas às unidades de contexto, originando os núcleos de sentido relacionados às categorias” (ANDRADE, 2013).

A Etapa 3 (cor azul da Figura 4.2) representa todos os procedimentos referente a rede semântica das 17 entrevistas. Calculamos todos os índices relevantes para a nossa análise, que são: 3.897 **vértices**, 120.047 **arestas**; $91 < k > m$ (**grau médio**); 0,016 **Densidade**,

¹Software que possibilita comunicações de voz e vídeo via Internet, permitindo a chamada gratuita entre usuários em qualquer parte do mundo.

5 **Diâmetro**, 2,286 **CMM**, 0,762 **CAM**, 0,095 **Modularidade** e ao toda a rede foi subdividida em 14 **Comunidades**. Além da aplicação da distribuição espacial das redes semânticas de cliques em comunidades. O resultado do DERC é representado na Figura 4.3:

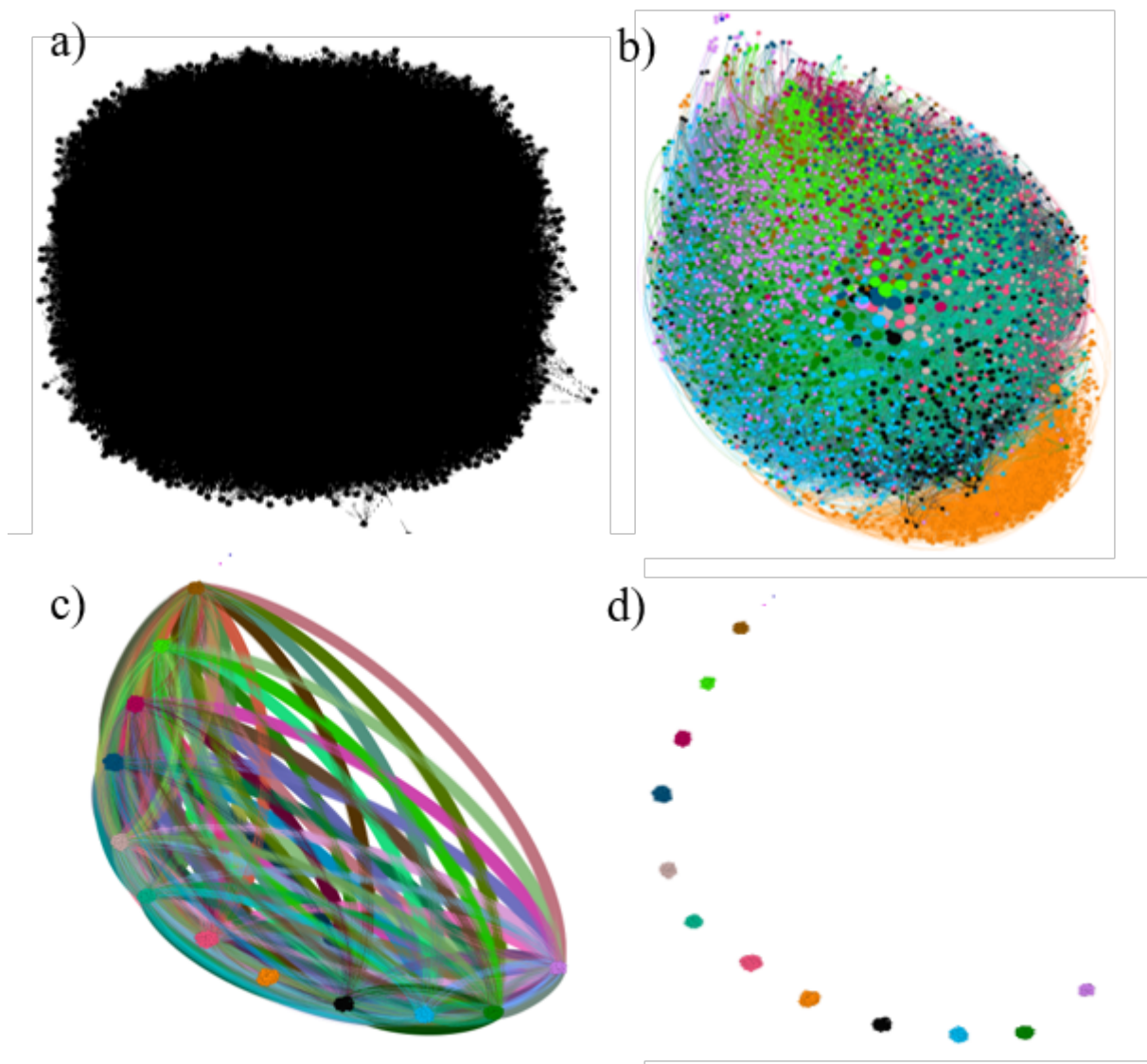


Figura 4.3: Rede semântica de todos os discursos do cenário 1. Fonte: Autora.

Na Etapa 4 (cor roxa na Figura 4.2) que consiste na finalização da sistematização dos dados na Análise de Conteúdo, ou seja, a Categorizaccão, além da sistematização dos dados da rede para posterior análise comparativa. Vale lembrar, que a autora [Andrade \(2013\)](#) elaborou as Categorias do seu discurso *a priori*, e o Modelo computacional de RSBC elaborou a *posteriori*.

A Etapa 5 (cor vermelha na Figura 4.2) representa a parte final que realiza uma análise comparativa entre os resultados da AC e os resultados das Redes Semânticas.

4.2.2 Simulação no Cenário 2

A nossa segunda amostragem é um relatório técnico sobre educação com 58 páginas. Aplicamos o nosso modelo para verificar se o nosso modelo computacional seria capaz de sugerir NS, UC e CAT para a Análise de Conteúdo.

4.2.3 Simulação segundo cenário

Como na primeira amostragem, aplicamos o nosso modelo computacional ilustrado na Figura 4.4

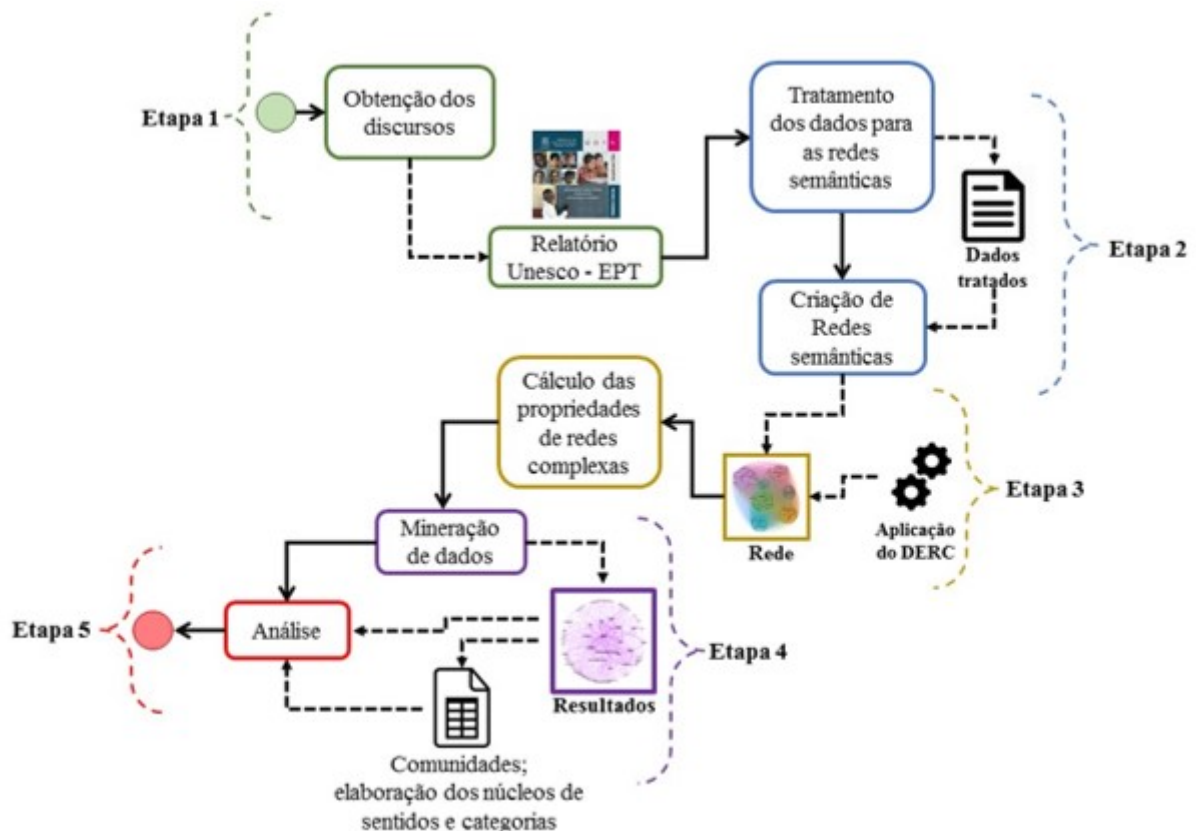


Figura 4.4: Modelo computacional aplicado no segundo cenário. Fonte: Autora.

Na Etapa 1 (cor verde na Figura 4.4) escolhemos o tema e obtivemos o discurso através

do site da UNESCO², o documento escolhido foi o Relatório Técnico de Educação - EPT 2015.

Na Etapa 2 (cor azul na Figura 4.4), o discurso passou pelos tratamentos manuais e computacionais descritos na Etapa 2 (cor azul da Figura 3.1).

Nesta Etapa 3 (cor ocre na Figura 4.4) calculamos os índices de redes (principalmente o *grau* dos vértices), aplicamos o algoritmo de modularidade (implementado no *Gephi 0.9.1*) e o DERC para melhorar a visualização da rede.

A Etapa 4 (cor roxa na Figura 4.4) é a verificação dos resultados, rede semântica do cenário II foi particionada em 7 comunidades, como ilustrado na Figura 4.5:

²<http://unesdoc.unesco.org/images/0023/002325/232565por.pdf>

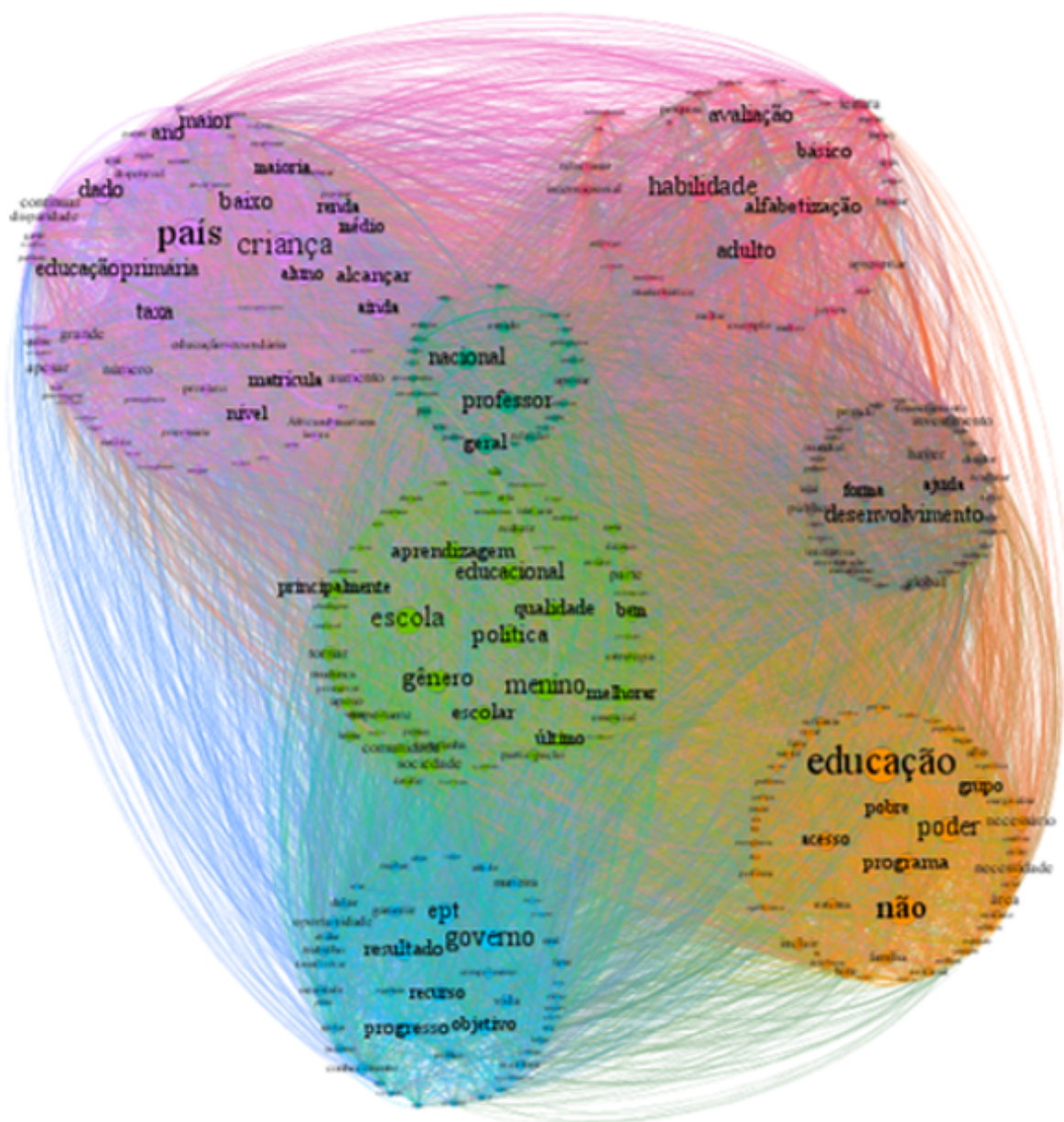


Figura 4.5: Rede geral com as 7 comunidades e suas ligações. Fonte: Autora.

Caso seja necessário também é possível visualizar as comunidades sem as arestas, como ilustrado na Figura 4.6:



Figura 4.7: Sentido na elaboração da autora Andrade: Categorias, Núcleos de Sentido e Unidade de Registro da autora [Andrade \(2013\)](#). Fonte: Autora.

Na nossa pesquisa, por falta de um algoritmo que pudesse melhor correlacionar com o processo ilustrado na Figura 4.7, decidimos elaborar as Categorias somente ao fim da análise das Unidades de Registro e elaboração dos Núcleos de Sentido, assim demonstrado na Figura 4.8



Figura 4.8: Sentido da nossa elaboração: das Unidades de Registro, Núcleos de Sentido e Categoria da nossa pesquisa. Fonte: Autora.

Posto isto, a primeira parte para da análise dos resultados da amostragem I começa quando analisamos as 10 primeiras palavras de maior frequência de cada um dos 7 Núcleos de Sentidos e comparamos com o grau de cada uma das palavras na rede, ilustrado na Tabela 4.1

Tabela 4.1: Comparação das 10 palavras de maior frequência do Núcleo de Sentido 1 (NS1: Difusão do Conhecimento - Categoria 1: Difusão do Conhecimento e Colaboração) com seu grau na Rede. Fonte: Autora.

CATEGORIA 1: Difusão do conhecimento e colaboração	Frequência	Grau
Núcleo de sentido 1: Difusão do conhecimento		
Pesquisa	172	1393
Trabalho	46	1291
Conhecimento	126	1250
Área	98	1190
Publicação	185	837
Difusão	62	593
Projeto	68	574
Artigo	68	533
Empecilho	65	411
Difusão do conhecimento	46	402

É possível percebermos que praticamente 100% das palavras escolhidas para compor os NS1 (NS1: Difusão do Conhecimento - Categoria 1: Difusão do Conhecimento e Colaboração) possuem um valor alto na propriedade de rede grau, ou seja, há uma consistência entre o que a RS aponta e as escolhas das palavras do Analista. A mesma exatidão dos 100% das palavras com alta frequência ocorrem no NS5 (NS5: Comunidade Acadêmica - Categoria: Estrutura Acadêmica), vejamos a Tabela 4.2:

Tabela 4.2: Comparação das 10 palavras de maior frequência do Núcleo de Sentido 5 (NS5: Comunidade Acadêmica - Categoria: Estrutura Acadêmica) com seu grau na Rede. Fonte: Autora.

NS5: Comunidade Acadêmica	Frequência	Grau
Pesquisador	205	1284
Pessoa	63	1275
Grupo	222	1245
Professor	47	657
Aluno	75	616
Grupo de Pesquisa	81	399
Doutorando	48	151
Programa de pós graduação	59	131
Comunidade científica	62	102

Reproduzimos essa comparação com os 7 Núcleos de Sentido descritos na Tabela 3.1, e encontramos outro fato relevante nesta comparação, os NS que possuem pouca expressão na análise da amostragem I, isto é, palavras com baixa Frequência, quase não são percebidas ou capturadas na rede, no caso do NS7 (Política de Publicação) apenas as palavras: “Critério” e “Qualidade” estão entre as 10 primeiras palavras com maior grau, entre as comunidades formadas na rede semânticas. Vejamos a Tabela 4.3:

Tabela 4.3: As 10 palavras que compõem o Núcleo de Sentido 7 (NS7: Política de Publicação - Categoria 3: Política de CTI). Fonte: Autora.

NS7: Política de Publicação	Frequência
Capes	44
Política	37
Qualis	28
Cnpq	24
Política de publicação	24
Currículo	19
Critério	18
Avaliação	17
Qualidade	13

4.4 Resultados cenário 2

Para à elaboração dos núcleos de sentido e categorias da segunda amostragem foi necessário estudar individualmente as comunidades. Assim sendo, selecionamos os 20 vértices de maior grau de cada comunidade identificada (um total de 140 vértices) e obtivemos, por exemplo, a Comunidade 2 que sugere a partir das suas conexões e grau dos vértices um NS sobre “Educação Básica”, ilustrado nas Figura 4.9:

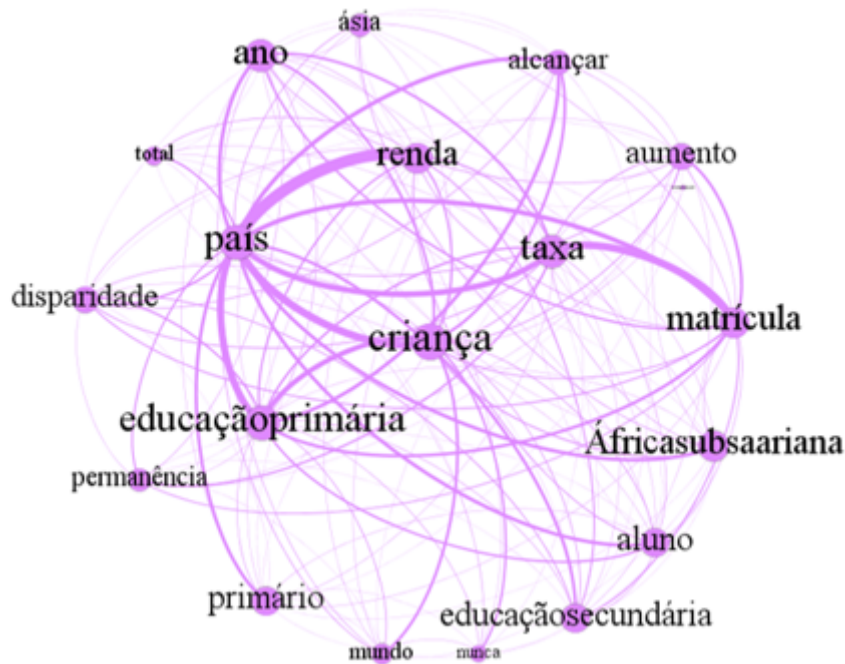


Figura 4.9: Comunidade 2 gerada pelo algoritmo de modularidade com suas 20 palavras com maiores vértices. Fonte: Autora.

Já Comunidade 6, seu tema central, quando analisamos suas 20 palavras com maior grau, indica um possível núcleo de sentido sobre “desigualdade social na educação”, demonstrado na Figura 4.10:

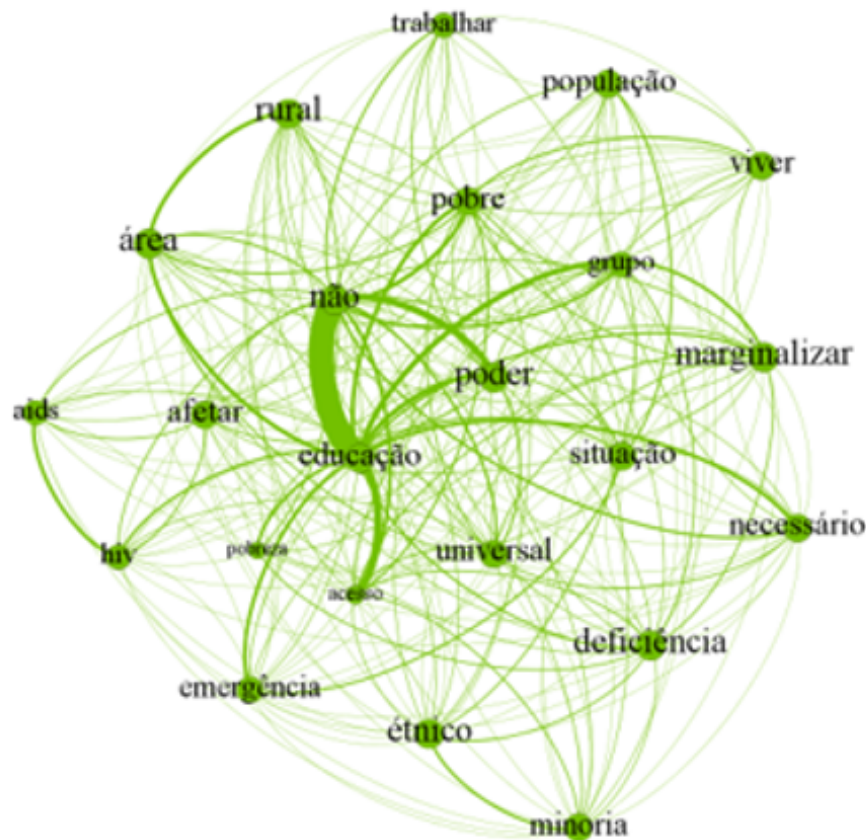


Figura 4.10: Comunidade 6 gerada pelo algoritmo de modularidade com suas 20 palavras com maiores vértices. Fonte: Autora.

Após as devidas análises em todas as comunidades, obtivemos como resultado 3 Categorias e 7 Núcleos de Sentido, ilustrado na Tabela 4.4:

Tabela 4.4: Tabela das CAT e NS a partir da aplicação do nosso modelo computacional. Fonte: Autora.

Categorias	Núcleo de sentido
Educação e Políticas públicas	Cuidados na primeira infância
	Qualidade da educação
	Mortalidade infantil
Educação primária universal	Alfabetização de jovens e adultos
Desigualdade na educação	Educação básica
	Disparidades de gênero na educação primária e secundária
	Acesso restrito à educação pobres

É perceptível, a partir dos exemplos supracitados, que para a elaboração de Núcleos de

Sentido e Categorias a *posteriori* o nosso modelo computacional é muito satisfatório, vale ressaltar que realizamos esta análise de conteúdo em um documento de 58 páginas. Ademais só foi possível a construção desses Núcleos de Sentidos e Categorias por conta da análise das propriedades das redes semânticas baseadas em cliques.

Foi possível perceber também a relação das *Cliques* e com as Unidades de Contexto. Pelo fato das *Cliques* se formarem a partir de sentenças, e as Unidades de Contexto serem também sentenças. Ademais o algoritmo de modularidade subdivide por afinidade a rede analisada em comunidades formadas por *Cliques*, que por sua vez o conjunto dessas *Cliques* sugestionam os Núcleos de Sentidos e possíveis Categorias.

Considerações Finais e Contribuições

Ao iniciarmos esta pesquisa tínhamos enquanto objetivo principal apoiar o método de Bardin por meio das redes semânticas baseadas em cliques para assim ser possível realizar a análise de AC em documentos extensos de forma menos laborosa e que demandasse menos tempo. Acreditamos que o nosso trabalho concluiu quatro pontos relevantes para o cumprimento dos objetivos.

Elaboramos um método para Distribuição espacial das redes em comunidades utilizando o *software Gephi* que possibilita a visualização mais clara da rede, auxiliando a interpretação do pesquisador ao analisar o documento, expondo as conexões/relações entre as palavras e/ou sentenças.

Ao realizar a AC com a adição do nosso modelo computacional, o pesquisador terá mais subsídios para sustentar as suas argumentações, pelo fato das redes semânticas baseadas em cliques serem outra maneira de representar o conhecimento de forma sistematizada, além de todo o seu embasamento teórico e análise serem pautados em propriedades matemáticas, quase eliminando discussões subjetivas, assim contribuindo para o rigor científico do método da AC.

A partir do nosso modelo computacional torna-se viável a análise de conteúdo em documentos extensos, pois as RS baseadas em cliques apontam de maneira simples as palavras com maior importância na rede através da propriedade *Grau*; proporciona a identificação de forma mais célere das UC por conta das cliques. Porque as cliques são as próprias sentenças (frases), ou seja, a própria UC da palavra vértice analisado; subdivide o discurso analisado em grupos por afinidades, assim sugerindo os NS para a construção de possíveis CAT.

Avançamos no uso de um método quantitativo na área de educação. Possibilitando que outros pesquisadores da área das Ciências Humanas se aventurem em utilizar os métodos quantitativos em suas pesquisas.

Sendo assim, a partir das similaridades entre os dois métodos, suas aplicações e auxílios, concluímos que é possível e viável utilizar as RS baseadas em cliques para apoiar o método da Análise de Conteúdo.

5.1 *Trabalhos futuros*

Pretendemos continuar na seara da pesquisa interdisciplinar, aprofundando nossos esforços e estudos sobre redes semânticas com ênfase nas comunidades. Em tempo, vale ressaltar, que já estamos aplicando nosso modelo em duas pesquisas distintas, que detalharemos a seguir.

A primeira pesquisa, em parceria com o pesquisador Pedro Nery da UnB - Universidade de Brasília, refere-se a Operação Lava Jato e seus investigados. Os nomes são conectos a partir da seguinte situação: processos em comum, isto é, os casos em que os vértices constam juntos como réu ou investigado ou requerido. Segue ilustrado na Figura 5.1

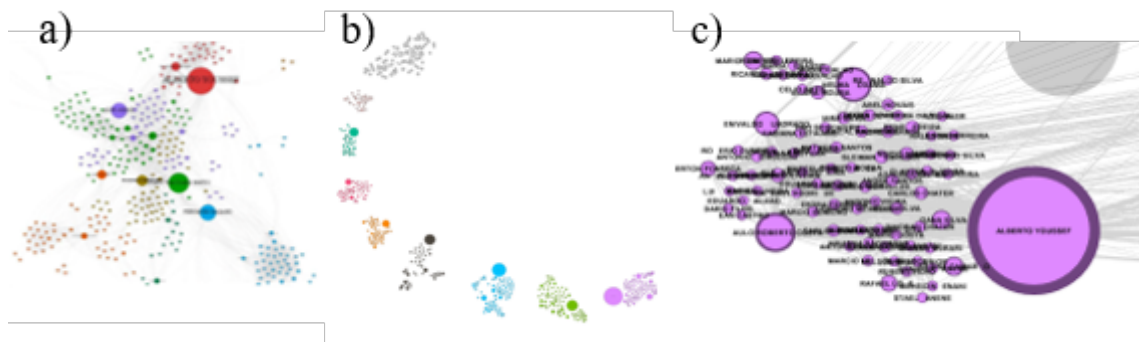


Figura 5.1: Rede dos processados, investigados ou réus da Lava Jato. Fonte: Autora.

O segundo caso refere-se a uma tese doutoral já concluída (NETO, 2016) que propôs um metamodelo explicativo das relações sistêmicas entre os indivíduos em grupos de ajuda-mútua de Narcóticos Anônimos no processo de recuperação do uso de drogas.

A Figura 5.2 retrata a rede semântica das respostas das entrevistas realizadas por Neto (2016) com os NA,

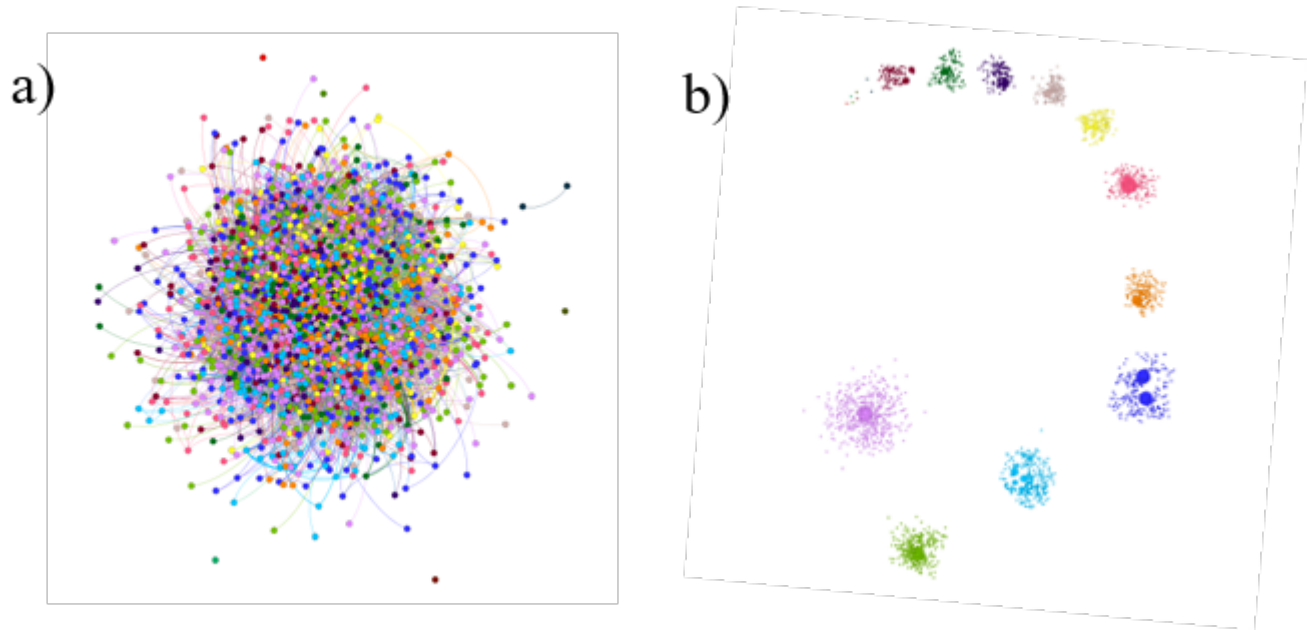


Figura 5.2: Rede Semântica das entrevistados da NA. Fonte: Autora.

Pode-se concluir que o modelo computacional para o apoio da Análise de conteúdo congrega uma série de otimizações que foram abordadas ao longo desta dissertação, especificamente neste Capítulo. Os resultados esperados para elaboração e identificação das UC, NS e CAT foram alcançados e comprovados neste trabalho.

Por fim, desejamos aprofundar mais os estudos sobre Redes Semânticas e comunidades, afim de aprimorar o modelo computacional proposto e subsidiar novas pesquisas em qualquer área do conhecimento.

Referências Bibliográficas

- ALBERT, B.; BARABÁSI, A. Topology of evolving networks: Local events and universality. *Physical Review E*, v. 85, n. 24, p. 52345237, 2000.
- ANDRADE, M. T. T. *A colaboração em comunidades científicas interdisciplinares: das redes de coparticipação à difusão do conhecimento*. Tese (DMMDC) — UFBA, Salvador, 2013.
- B. JEAN-L. G., R. L. V. D.; L., E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment*, v. 08, n. P10008, p. 1742–5468, 2008.
- BARABÁSI, A. *Linked (conectado): a nova ciência dos networks*. São Paulo: Leopardo Editora, 2009.
- BARABÁSI, A. *Network Science*. United Kingdom: Cambridge, 2016.
- BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 1977.
- BARDIN, L. *Análise de conteúdo*. Lisboa: Edições Setenta, 2010.
- BARRETO, R. S. F. D. et al. A teoria de redes como subsídio potencializador para a análise de conteúdo. *SIINTEC*, p. 94–102, 2016.
- CALDEIRA, S. Mestrado em Modelagem Computacional, *Caracterização da rede de signos linguísticos: um modelo baseado no aparelho psíquico de freud*. Salvador: [s.n.], 2005.
- CUNHA, F. J. A. P. *Da adesão à participação em uma rede de hospitais como promoção da aprendizagem organizacional e da inovação gerencial: um olhar sobre a rede inovar-ha*. Tese (Faculdade de Educação) — UFBA, Salvador, 2012.
- EULER, L. Solutio problematis ad geometriam situs pertinentis. *Commentari Academiae Scientiarum Imperialis Petropolitanae*, v. 8, p. 128140, 1736.
- FADIGAS, I.; I.; TRAZÍBULO, H.; MORET, M.; PEREIRA, H. B. B. Análise de redes semânticas baseada em títulos de artigos de periódicos científicos: o caso dos periódicos de divulgação matemática. *Educ. Mat. Pesqui*, v. 11, n. 1, p. 167–193, 2009.
- FADIGAS, I.; PEREIRA, H. B. B. A network approach based on cliques. *Physica A: Statistical Mechanics and its Applications*, v. 392, n. 10, p. 2576–2587, 2013.
- GOLDENBERG, M. *A arte da pesquisa - Como fazer pesquisa qualitativa em Ciências Sociais*. Rio de Janeiro: Editora Record, 2001.

- GRILO, M. et al. Robustness in semantic networks based on cliques. *Physical A*, v. 472, p. 94–102, 2017.
- GROSS, J.; YELLEN, J. *Graph theory and its applications*. Boca Raton: CRC Press, 1999.
- NETO, J. L. A. L. *Metamodelo explicativo das relações sistêmicas entre os indivíduos em grupos de ajuda-mútua de Narcóticos Anônimos*. Tese (UFBA) — UFBA, Salvador, 2016.
- NEWMAN, M. E. J. The structure and function of complex networks. *Siam Review*, v. 250, n. 2, p. 167256, 2003.
- PAIXAO, R. Mestrado em Modelagem Computacional e Tecnologia Industrial, *Visualização de Informação em Redes Sociais e Complexas Utilizando o Guará Script*. Salvador: [s.n.], 2016.
- PEREIRA, H. B. B.; FADIGAS, I.; SENNA, V.; MORET, M. Semantic networks based on titles of scientific papers. *Physica A: Statistical Mechanics and its Applications*, v. 360, n. 6, p. 11921197, 2011.
- STERNBERG, R. *Psicologia Cognitiva*. Porto Alegre: Artes Médicas Sul, 2011.
- STROGATZ, S. H. Exploring complex networks. *Nature*, v. 410, p. 268276, 2001.
- TEIXEIRA, G. Mestrado em Modelagem Computacional, *Redes semânticas em discursos orais: uma proposta metodológica baseada na psicologia cognitiva utilizando redes complexas*. Salvador: [s.n.], 2007.
- UNESCO. *Educação para todos 2000-2015: progressos e desafios*. Place de Fontenoy: UNESCO, 2015.
- WATTS, D. D.; STROGATAZ, S. H. Collective dynamics of smallworld networks. *Nature*, v. 393, n. 4, p. 440442, 1998.
- WATTS, D. J. *Small World*. Princeton: Princeton University Press, 1999.

Redes semânticas baseada em cliques para apoio à análise de conteúdo

Renata Souza Freitas Dantas Barreto

Salvador, Fevereiro de 2018.