# Análise da estrutura da rede de frames da FrameNet Brasil

**Míria Bóbó[1], Victor Ströele[1], Ely Edison da Silva Matos[2], Regina Braga[1], Fernanda Campos[1], José Maria N. David[1], Tiago Timponi Torrent[2]**

[1]Departamento da Ciência da Computação – Universidade Federal de Juiz de Fora (UFJF)
Caixa Postal 20.006 – 36.016-970 – Juiz de Fora – MG – Brazil

[2]FrameNet Brasil – Universidade Federal de Juiz de Fora (UFJF)
Caixa Postal 20.006 – 36.016-970 – Juiz de Fora – MG – Brazil

{miria.luisa,victor.stroele,jose.nazar}@ice.ufjf.br,

{ely.matos,regina.braga,fernanda.campos,tiago.torrent}@ufjf.edu.br

***Abstract.*** *FrameNet Brazil is a computational lexicography project that aims to use frames for the description of the meaning of the word. It generates a network of frames that bind to specific relationships. The process of creating or updating a frame can pass through more than one Linguistic expert at different times, which can generate a change in the network structure that does not reflect the conceptual model of the project. The objective of the work is to make an analysis of the frame network structure of FrameNet Brazil and to identify critical points that may reflect in violations of the constraints of the its conceptual model. The results found were analyzed by a specialist in the field of Linguistics and his considerations indicate that the proposed solution is viable in the analysis of the network of FrameNet Brazil.*

***Resumo.*** *A FrameNet Brasil é um projeto de lexicografia computacional que tem como objetivo usar frames para a descrição de significados de palavras. Ela gera uma rede de frames que se ligam com relacionamentos específicos. O processo de criação ou atualização de um frame pode passar por mais de um especialista da Linguística, em momentos diferentes, o que pode gerar uma mudança na estrutura da rede que não reflete o modelo conceitual do projeto. O objetivo do trabalho é fazer uma análise da estrutura da rede de frames da FrameNet Brasil e identificar pontos críticos que possam refletir em violações de restrições do modelo conceitual da mesma. Os resultados encontrados foram analisados por um especialista da área da Linguística e as suas considerações indicam que a solução proposta é viável no que se refere à análise da rede da FrameNet Brasil.*

## 1. Introdução

Redes complexas podem ser definidas como grafos que apresentam topologias complexas [Barabási et al. 2000]. Elas podem ser usadas como abstrações para representar situações do mundo real onde existam relacionamentos entre pares de entidades envolvidas, como por exemplo palavras de uma frase e as ligações entre elas.

A FrameNet Brasil[1] é um laboratório de Linguística Computacional, cuja missão é desenvolver soluções computacionais para problemas de Processamento de Linguagem

[1]http://www.ufjf.br/framenetbr/

Natural, usando Semântica de *frames*. O projeto central do laboratório é a manutenção da FrameNet[2], que é uma rede complexa formada por *frames* - estruturas cognitivas que definem situações, objetos ou eventos através de cenários - e seus relacionamentos.

A criação de *frames* é feita por linguistas, através de softwares específicos. Contudo, devido a características inerentes ao projeto, um *frame* pode ser atualizado por diferentes especialistas. As mudanças efetuadas podem violar as restrições do modelo conceitual da FrameNet, visto que parte do processo de criação e/ou atualização dos *frames* apela para a intuição do linguista [Fillmore et al. 2003b]. Assim sendo, este artigo se propõe a fazer uma análise da estrutura da rede de *frames*, com o objetivo de identificar situações que possam refletir em violações de restrições do modelo conceitual da FrameNet Brasil. As características identificadas foram avaliadas por um especialista como uma forma de avaliação da solução proposta.

Como contribuições deste trabalho podemos destacar: (i) a concepção de uma abordagem para análise da FrameNet Brasil, possibilitando a identificação de pontos críticos que devem ser analisados pelos especialistas; (ii) a caracterização dos elementos dessa rede; e (iii) o uso de métricas de redes complexas para extrair informações sobre a estrutura da rede.

O restante do artigo está organizado da seguinte forma: na seção 2 são contextualizados o projeto FrameNet bem como os conceitos associados a ele. O Modelo Conceitual é descrito na seção 3 e a proposta do artigo na seção 4. As considerações finais e trabalhos futuros são apresentados na seção 5.

## 2. Referencial Teórico

Nesta seção é apresentada a FrameNet, projeto precursor da FrameNet Brasil, e são definidos os conceitos necessários para o entendimento deste artigo.

### 2.1. FrameNet

A FrameNet, também chamada de FameNet de Berkeley, é um

> "projeto de lexicografia computacional que extrai informações sobre links semânticos e sintáticos vinculados às palavras em inglês de grandes corpus de textos eletrônicos, usando procedimentos manuais e automáticos, e apresenta esta informação em uma variedade de relatórios baseados na web" [Fillmore et al. 2003a, p. 1].

Ela foi criada em 1997, liderada por Charles J. Fillmore, e surgiu do cruzamento da Semântica de *Frame* com Lexicografia (geração de dicionário). Atualmente ela foi expandida para outras línguas como Espanhol, Alemão, Sueco, Letão, Japonês, Chinês, Coreano e Português brasileiro.

O projeto tem como objetivo organizar as descrições lexicográficas por *frames* e usar os dados extraídos de corpus para descobrir todas as funções semânticas e propriedades gramaticais das palavras - *unidades lexicais* - que evocam o mesmo *frame* [Salomao 2009].

A Semântica de *Frames*, ou semântica da compreensão, foi gerada como uma abordagem para solucionar problemas da semântica lexical (dar significados as

[2]http://webtool.framenetbr.ufjf.br/index.php/fnbr/report/frame/main

palavras) [Salomao 2009]. A ideia central é que os significados das palavras devem ser descritos em relação aos *frames* semânticos, que são "representações esquemáticas das estruturas conceituais e padrões de crenças, práticas, instituições, imagens, etc., que fornecem uma base para uma interação significativa em uma determinada comunidade de fala" [Fillmore et al. 2003a]. Neste sentido, os *frames* são pacotes de conhecimento que moldam e permitem que os humanos deem sentido às suas experiências [Fillmore and Baker 2010]. Os *frames* que são evocados se baseiam no conhecimento que temos sobre os fenômenos e sua associação com os valores culturais. Um exemplo é descrito a seguir:

*Mary foi convidada para a festa de Jack. Ela se perguntou se ele gostaria de um brinquedo.*

Os *frames* evocados na frase acima estão ancorados ao verbo `convidar` - que indica um relacionamento envolvendo um anfitrião, um convidado e uma ocasião - e ao substantivo `festa` - que evoca um evento social geralmente com um anfitrião, convidados e uma ocasião. O trecho *festa de Jack* remete a uma festa na qual o Jack é o anfitrião ou na qual Jack é celebrado. Não se verificam implicações linguísticas que evoquem diretamente o *frame* `festa de aniversário` porém o substantivo *brinquedo*, a preocupação *se o Jack iria gostar de um* e os outros detalhes fornecidos pela linguagem, permitem que o leitor infira que o Jack é o aniversariante, que brinquedo é o presente de aniversário, que a Mary é a convidada e assim por diante.

### 2.2. Estrutura semântica da FrameNet

A FrameNet é constituída por *unidades lexicais*, *frames*, *elementos de frame* e *relacionamentos*. Nesta seção serão apresentados, de forma detalhada, cada um destes elementos, que são usados na descrição da rede formada pela FrameNet Brasil.

i) *Unidade Lexical (UL)*
Uma *Unidade Lexical - UL* é a palavra quando lhe é atribuída um dos seus significados, ou seja, é um emparelhamento de uma palavra com um significado que pertence a um *frame* [Ruppenhofer et al. 2016]. Uma palavra com quatro significados é tratada como quatro unidades lexicais e, na maioria dos casos, ela pode pertencer a mais de um *frame* [Fillmore et al. 2004]. Dizemos que a palavra *evoca* um *frame* quando o significado dela é baseado no *frame*.
A meta é que toda *UL* evoque um *frame* porém, ela deve destacar algum elemento desse *frame* de forma particular [Salomao 2009]. Por exemplo, o *frame* `Aplicar-Calor` é o que descreve uma situação envolvendo cozinha, comida e um instrumento de aquecimento, e ele é evocado pelas palavras *assar, cozer, ferver, secar, borbulhar, corar, dourar, grelhar, vapor, etc*. Essas palavras são chamadas de *Unidade Lexicais*.

ii) *Frames*
Um *Frame* é uma "estrutura conceitual que descreve um tipo particular de situação, objeto ou evento, juntamente com seus participantes e adereços" [Ruppenhofer et al. 2016]. É um sistema de conceitos relacionados de modo que para compreender qualquer um deles, é necessário compreender o sistema como um todo. Ele é composto por elementos que ajudam a completar o seu significado

(*Elementos de Frame*), evocado por uma *Unidade Lexical* e se liga a outros *frames* com relacionamentos específicos.

iii) *Elementos de Frames (EF)*
Os *Elementos de Frame - EF* são atributos usados como etiquetas para as palavras ou frases que estão na construção gramatical com as *UL* que evocam o *frame* [Fillmore et al. 2004]. São os papéis semânticos das entidades envolvidas em cada *frame* [Fillmore et al. 2003a].
Os *EF* existem na estrutura do *frame* porém, podem ou não estar representados na frase em que o *frame* é evocado. Por exemplo o *frame* `Danificar` - que é definido como um *Agente* que afeta um *Paciente* de modo que este mude para um estado não-canônico - possui como principais *EFs* os atributos:

- *Agente:* A entidade consciente, geralmente uma pessoa, que realiza a ação intencional que resulta no dano ao *Paciente*;
- *Paciente:* A entidade que é afetada pelo *Agente*, para que esteja danificada;
- *Causa:* Um evento que leva ao dano do *Paciente*.

iv) *Tipos Semânticos*
O principal objetivo de usar *Tipos Semânticos* na FrameNet é de demonstrar informações que não são bem representadas na hierarquia de *frames* [Fillmore et al. 2003a]. Eles também podem ser usados para expressar importantes diferenças semânticas entre *ULs* que se repetem em vários *frames* [Ruppenhofer et al. 2016]. Por exemplo, as *ULs* de um *frame* podem possuir avaliações positivas - *louvar* do *frame* `Julgamento`, *gostar* do *frame* `Sujeito_experimentador` - ou negativas - *criticar* do *frame* `Julgamento`, *detestar* do *frame* `Sujeito_experimentador`.

Atualmente a FrameNet cobre mais de 13.000 *Unidades Lexicais*, distribuídas em mais de 1.200 *Frames* e atestadas por mais de 200.000 frases anotadas.

### 2.3. Relacionamentos

A FrameNet é uma rede de *frames* em que as ligações ocorrem através de relações específicas (*Herança, Perspectiva, Uso, Subframe e Precedência*). As relações são usadas para melhorar a compreensão dos *frames* e proporcionar robustez (visto que *frames* semanticamente similares podem estar associados, apesar de estarem separados) [Ruppenhofer et al. 2016]. A Figura 1 ilustra relacionamentos entre *frames* e seguir são descritos cada um deles.

i) *Herança*
É o relacionamento mais forte na FrameNet [Ruppenhofer et al. 2016]. Ocorre entre um *frame* pai e um *frame* filho, onde *frame* filho herda todos ou parte dos *EFs*, *subframes* e tipos semânticos do pai. Os *EFs* do *frame* filho não têm, necessariamente, os mesmos nomes dos *EFs* do pai e podem ser adicionados outros, dadas as especificidades do *frame* filho. Por exemplo, o *frame* `Fornecer` herda do *frame* `Dar` e, além de possuir os *EFs Tema* e *Destinatário*, ele também especifica que o doador é um *Fornecedor* e que tem que haver o *Objetivo* do tema proposto.

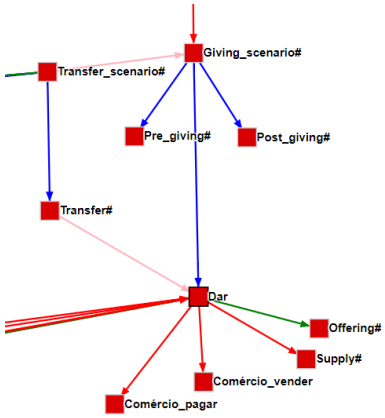



**Figura 1. Relacionamentos do *frame* `Dar` com os outros *frames***

ii) *Uso*
É a relação usada quando parte do cenário evocado pelo *frame* filho se refere a um *frame* pai. É um tipo de relacionamento de herança, onde o *frame* filho pode usar (herdar de) múltiplos *frames* pai. Por exemplo, o *frame* `Oferta` usa o *frame* `Dar`, já que o *Ofertante* oferece um *Tema* para um *Potencial_Destinatário* e somente quando o *Potencial_Destinatário* aceita a oferta é que a `Transferência` ocorre.

iii) *Perspectiva*
Esta relação é semelhante a de *Uso* e consiste em indicar, pelo menos, dois pontos de vista de um *frame* neutro. Por exemplo, o *frame* `Dar` é uma perspectiva do *frame* `Transferência`, uma vez que os verbos *dar* e *receber* expressam o ponto de vantagem da cena (do *Doador* ou do *Destinatário* respectivamente).

iv) *Subframe*
Essa relação é usada para representar sub-eventos ou subpartes de um *frame* complexo. Os *subframes* comumente se referem à sequência de passos ou transações e podem ser descritos, separadamente, por *frames*. Por exemplo, o *frame* `Dar` juntamente com os *frames* `Pré_Doação` e `Pós_Doação` são *subframes* do *frame* `Cenário_Doação`.

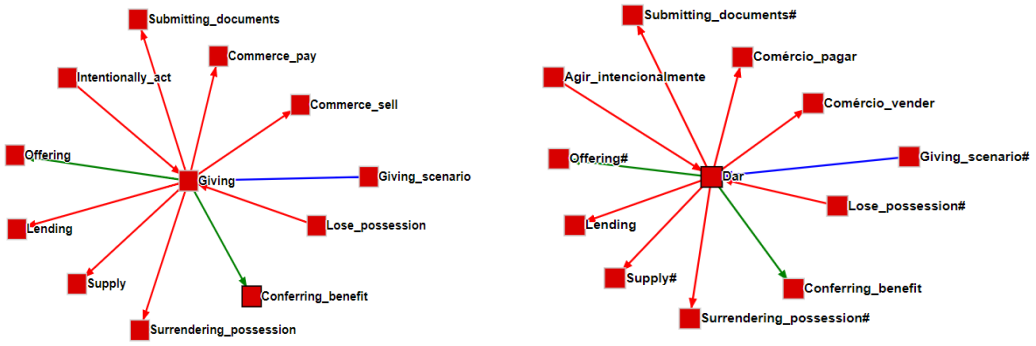
v) *Precedência*
Essa relação ocorre apenas entre dois *subframes* de um *frame* complexo para especificar a sequência dos eventos de um certo cenário. Este é o único relacionamento que pode formar ciclos. No *frame* `Cenário_Doação`, o relacionamento *Precedência* é usado para conectar o *frame* `Pré_Doação` ao *frame* `Dar` e o *frame* `Dar` ao *frame* `Pós_Doação`, pois eles devem ocorrer em ordem.

### 2.4. FrameNet Brasil

A extensão do projeto inicial para outras línguas inclui o uso do estrutura definida no Inglês - *Unidades Lexicais, Frames, Elementos de Frames e suas conexões* - além da implementação de novas funcionalidades. A FrameNet Brasil focou no desenvolvimento de um léxico, de um *Construction*, de um banco de dados trilíngue para o domínio do

esporte e turismo, e de uma FrameNet Brasil WebTool.

a) Léxico
O Léxico é uma expansão da FrameNet e possui uma rede que relaciona *frames* a *unidades lexicais*. Alguns *frames* não foram alterados no processo de expansão da FrameNet para a FrameNet Brasil, como é o caso do *frame* `Giving` que em português pode ser evocado pela *unidade lexical* `Dar`, onde seus *elementos de frame* (entre outras características) se mantiveram intactos. A Figura 2 mostra o mesmo *frame* evocado pelas *unidades lexicais* `Giving` e `Dar`.




(a) Unidade Lexical `Giving` (b) Unidade Lexical `DAR`

**Figura 2. Frame `Giving`**

b) *Construction*
O *Construction* é um repositório de construções gramaticais do Português brasileiro. As construções se assemelham a estrutura dos *frames*, contendo uma definição em prosa e um conjunto de *Elementos de Construção* que são os componentes da construção. A principal contribuição do *Construction* da FrameNet Brasil foi a criação de dois tipos de relacionamentos e 5 tipos de restrições na base de dados. Foi gerado o relacionamento *Herança* entre os construtores e o relacionamento *Evoca*, que liga o construtor ao *frame* que ele evoca.

c) Base de dados trilíngue
Uma FrameNet de domínio específico multilingual foi desenvolvida, como prova de que *frames* podem ser usados como representações interlinguais em um dicionário eletrônico trilíngue (português-inglês-espanhol) para o Turismo e para Copa do Mundo de Futebol [Torrent et al. 2014]. O dicionário (`http://framenetbrazildictionary.com`) possui 128 *frames* trilíngues, 1.125 *unidades lexicais* e mais de 13.000 frases anotadas gerando, automaticamente, uma lista de tradução equivalente para todos os verbos e substantivos que indicavam eventos . A Figura 3 mostra a interface do aplicativo do dicionário (`http://www.dicionariodacopa.com.br/`).

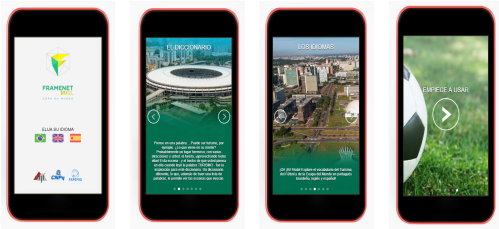


**Figura 3. Telas do aplicativo dicionário trilíngue**

d) FrameNet Brasil WebTool
O WebTool é o sistema de gerenciamento e anotação do banco de dados. Usa um banco de dados relacional, atualmente implementado no MySQL, que mantém os mesmos conceitos e estrutura usados na FrameNet, para facilitar a migração e o alinhamento de dados.

## 3. Modelo conceitual da FrameNet Brasil

A FrameNet procura capturar os *insights* humanos em estruturas semânticas de forma eficiente [Baker et al. 1998]. O processo começa com a caracterização do *frame* que a *UL* evoca, definindo o tipo de entidade, situação ou objeto que ele representa, escolhendo os rótulos dos *EFs* e a lista de *ULs* vinculadas a ele [Fillmore et al. 2003b].

Quatro etapas de processamento foram necessárias produzir o banco de dados da FrameNet Brasil:

- *Preparação:* descrições iniciais dos *frames, ULs* e *EFs* e verificação do padrão sintático de cada um, para o uso em consultas do subcorpus e anotação [Baker et al. 1998];
- *Extração de subcorpus:* geração de boas frases de exemplos através de ferramentas da linguística computacional;
- *Anotação:* marcação (a mão) dos *EFs* detectados no subcorpus e identificação de padrões de exemplos e frases com problemas;
- *Escrita da Entrada:* adição dos dados às tabelas do banco.

Os anotadores humanos são os que escolhem os termos que farão parte da lista de *ULs*, examinam o uso dos *EFs* e determinam os contextos sintáticos e colocacionais do significado do *frame* [Fillmore et al. 2003a] consultando dicionários eletrônicos e tesauros. Eles podem, em qualquer etapa de processamento, modificar uma decisão anterior (tomadas ou não por eles) com base em evidências do corpus e continuar o processo a partir desse ponto [Fillmore et al. 2003b]. Essas mudanças também podem implicar na alteração de *frames* já existentes e seus relacionamentos, podendo gerar um *loop* de alterações.

As relações entre os *frames* são implementadas no banco de dados, contendo as tabelas *frame, ULs, EFs* e *relacionamento* [Fillmore et al. 2004]. As tabelas e as relações entre elas procuram refletir a base teórica do projeto [Fillmore et al. 2003a].

## 4. Análise da Rede de *frames* da FrameNet Brasil

Nesta seção é apresentada a estrutura da rede de *frames* da FrameNet Brasil e as métricas de análise em redes complexas, com o intuito de mostrar como essa análise se reflete no

modelo conceitual da rede. Usou-se o software *Gephi*[3] para análise e visualização da rede e o software *yEd*[4] para a geração das árvores de *frames* (que serão descritas na seção 4.4).

### 4.1. Caracterização da rede de *frames*

A rede de *frames* é um grafo onde os vértices são os *frames* e as arestas são os relacionamentos entre eles. Possui 1359 nós e 1960 arestas. O grafo é dirigido pois os relacionamentos indicam uma certa hierarquia na rede. Existem *frames* mais genéricos que possuem muitos outros *frames* conectados a eles, gerando assim núcleos ou *clusters* bem definidos no grafo. Esses mesmos *frames* podem ser observados como pais das árvores geradas por suas conexões, onde quanto maior o nível do nó na árvore mais específico é o *frame*. Assim sendo, o grafo da FrameNet Brasil pode ser considerado como uma floresta de *frames* (vide seção 4.4) em que: os pais das árvores descrevem cenários mais gerais, os nós folha cenários mais específicos, os vértices podem se relacionar como irmãos e a altura da árvore pode ser um indicativo de quão bem estudado (aprofundado) foi o assunto em questão. A Figura 4 ilustra o grafo da FrameNet Brasil.

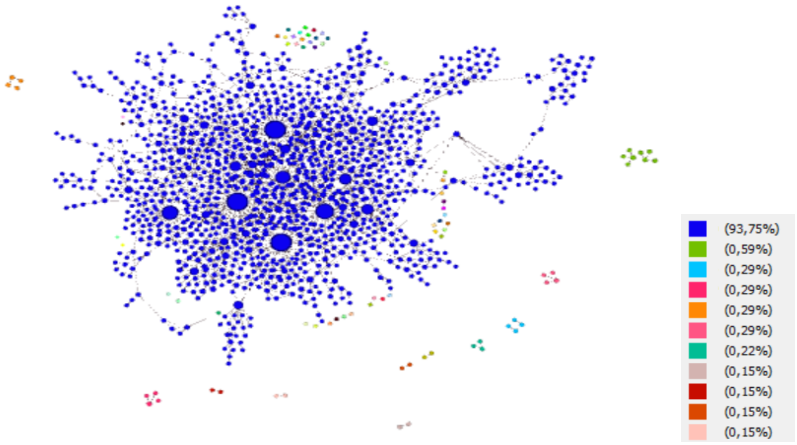



**Figura 4. Grafo da rede de *frames* da FrameNet Brasil. O tamanho do vértice foi determinado pelo seu grau. As cores assinalam as componentes conexas.**

### 4.2. Medidas de Centralidade

Foram aplicadas medidas de centralidade com o objetivo de detectar vértices de maior importância no grafo. A Tabela 1 lista essas medidas e o intervalo em que seus valores ocorrem.

**Tabela 1. Medidas de centralidade e o intervalo de valores. O *Closeness, Betweenness* e o *Eigenvector* possuem os valores contidos entre [0,1]**

| Intervalo | Grau | G. Entrada | G. Saída | Closeness | Betweenness | Eigenvector |
|---|---|---|---|---|---|---|
| Menor | 0 | 0 | 0 | 0.0 | 0.0 | 0.0 |
| Maior | 54 | 6 | 53 | 1.0 | 0.001601 | 1.0 |

[3]https://gephi.org/
[4]https://www.yworks.com/products/yed

O grau do nó indica a quantidade de relacionamentos que o vértice possui. Por ser um grafo dirigido, o valor do grau é a soma do grau de entrada e o de saída. O grau de saída foi usado para revelar vértices que poderiam apresentar um certa relevância no grafo, por possuírem valores maiores que o grau de entrada (como pode ser verificado na Tabela 1). Os vértices com maior grau de saída são os que possuem muitas relações de *herança* (por representarem cenários mais genéricos) e que, por essa razão, podem ser considerados como vértices críticos, visto que a sua remoção implicaria na fragmentação da árvore de *frames*, bem como do grafo. Foram identificados tanto vértices com grau (de entrada e de saída) nulo quanto vértices com grau de entrada maior que 1, e esses casos serão mais explorados adiante.

O *Closeness* foi usado para identificar os vértices mais centrais do grafo porém, os vértices com *Closeness* igual à 1 são os que possuem grau menor que 9; ou seja, são os vértices periféricos ou do penúltimo nível das árvores (pais das folhas). Isso ocorre por que o *Closeness* é uma medida que calcula a proximidade do vértice em relação aos outros que ele alcança, e nessa camada da rede os vértices alcançados são apenas os adjacentes.

O *Betweenness* foi usado para detectar os vértices *ponte* do grafo. O intervalo superior dessa medida é menor que 0,1. Isso acontece por que o grafo da FrameNet Brasil tende a uma estrutura de floresta de *frames* existindo poucos caminhos curtos que passam pelo mesmo vértice.

O *Eigenvector* foi usado para reconhecer vértices importantes com base na importância de seus vértices adjacentes, isto é, identificar vértices adjacentes a mais de um nó central. Contudo, como foi constatado anteriormente, os nós mais centrais deste grafo são os nós periféricos e, por esta razão, apenas um vértice apresentou o *Eigenvector* igual à 1.

### 4.3. Conectividade da rede

Visto que as medidas de centralidade não permitiram obter uma visão clara da estrutura da rede, optou-se por aplicar dois algoritmos de agrupamento (detecção de Componentes Conexas e de Comunidades) fornecidos pelo *Gephi*.

Para detecção de componentes conexas aplicou-se o algoritmo *Componentes Conexas* [Wasserman and Faust 1994]. Foram identificadas 60 componentes conexas, sendo que a maior engloba mais de 90% dos vértices (1274 *frames*). Nessa componente se encontram os vértices com maior grau e as árvores de *frames* mais altas. Das 59 componentes restantes 48 possuem apenas um vértice, isto é, existem vértices sem arestas no grafo o que indica que alguns *frames* estão sem relacionamentos na FrameNet Brasil. Estes *frames*, que representam 3% do total, deveriam ser analisados por especialistas para verificar se essa peculiaridade é aceita ou não semanticamente. A Figura 4 ilustra as componentes do grafo, onde as cores distinguem os vértices pertencentes a mesma componente.

Para detecção de comunidades aplicou-se o algoritmo de *Modularidade*, que consiste em agrupar os vértices de acordo com o valor da modularidade da partição [Blondel et al. 2008] (usou-se o valor de resolução=10 como parâmetro de entrada do algoritmo no *Gephi*). Foram detectadas 82 comunidades no grafo. Estas comunidades são grupos com ligações muito densas entres os vértices participantes. Os dois maiores grupos contêm dois dos vértices de maior grau e possuem mais de 90 membros, respectivamente.

Como pode ser observado na Figura 5, os vértices desconexos constituem também comunidades compostas de um único elemento.

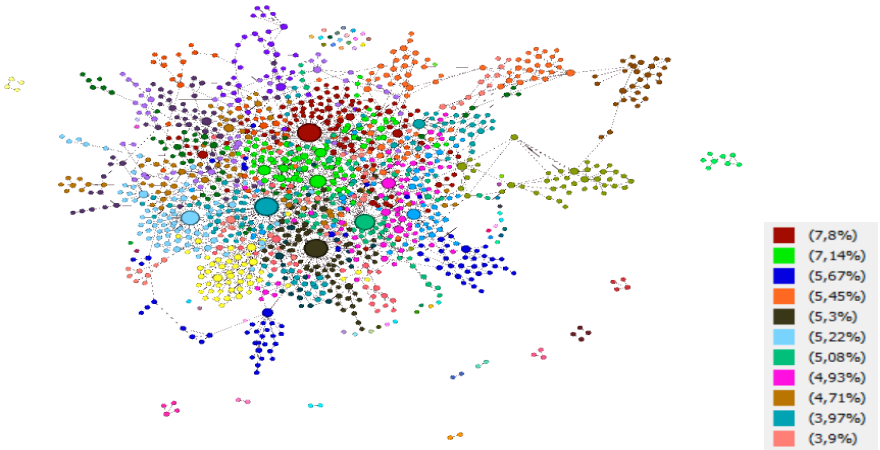
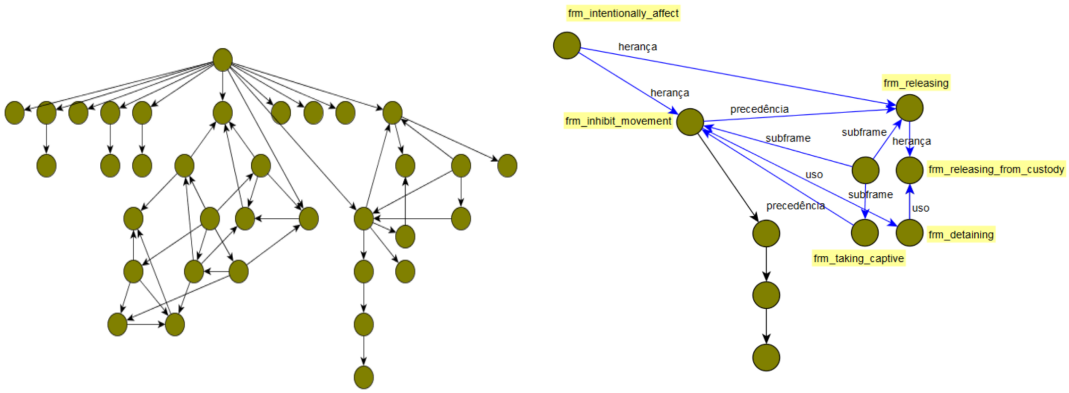



**Figura 5. Comunidades do grafo**

### 4.4. Árvores de *frames*

As ligações entre os *frames* que compõem um cenário geram uma estrutura hierárquica similar a de uma árvore, onde o nó pai é o *frame* mais geral e o nó folha o mais específico. A Figura 6(a) mostra a árvore gerada por uma das componentes conexas do grafo. Computacionalmente, os relacionamentos entre os *frames* violam as restrições da árvore (como a estrutura de dados) porém, semanticamente essas infrações podem fazer sentido para a descrição do assunto em questão (Figura 6(b)). Cabe a um especialista analisar e validar tais situações ou corrigir caso necessário.




(a) Árvore

(b) Relacionamentos entre os nós de uma das subárvore

**Figura 6. Visualização em árvore de *frames* de uma das componentes conexas do grafo**

### 4.5. Avaliação da proposta

As medidas e algoritmos apresentados nas seções anteriores foram empregados com o objetivo de identificar possíveis pontos críticos da rede. Foram detectados alguns casos,

a nível estrutural, e os mesmos foram encaminhados a um especialista em linguística computacional, para as respectivas avaliações. Doutor em Linguística Computacional, o especialista atua no projeto da FrameNet Brasil. Alguns casos de criticidade foram descartados, instantaneamente, e outros mereceram uma análise semântica mais profunda. Abaixo são descritos os casos e suas respectivas avaliações e/ou validações.

i) Medidas de centralidade
O grau do nó se revelou uma métrica válida para detecção de nós relevantes tanto a nível topológico quanto a nível semântico da rede. Nós com graus altos, quando eliminados da rede, tendem a gerar mais componentes conexas e o mesmo ocorre semanticamente, fazendo com que os *frames* adjacentes percam uma parte da informação associada a eles. Por exemplo, se o *frame* `Dar` for retirado da rede todos os *frames* adjacentes a ele perderão parte do significado (Figura 2). Esta medida foi validada de forma instantânea, contudo, o emprego do *Closeness, Betweennes* e *Eigenvector*, que são medidas que revelam o privilégio do nó em relação à sua localização, não se mostrou tão proveitoso. Os nós com *Closeness, Betweennes* e *Eigenvector* altos são, na maioria dos casos, nós periféricos da rede. Por serem casos atípicos à realidade computacional, essas medidas demandaram uma análise semântica mais acentuada. Todavia, foram descartadas por não oferecerem risco semântico à FrameNet.

ii) Conectividade
O algoritmo de detecção de componentes conexas evidenciou grupos pequenos de *frames* desconexos do restante da rede e até mesmo um número considerável de *frames* soltos (41), sem nenhum tipo de relacionamento. Quando apresentados, o especialista afirmou que estes casos podem ocorrer se o *frame* descrever um cenário muito específico. Todavia, essa subparte da rede não deixa de ser um motivo de alerta e receberá especial atenção dos linguistas do projeto.

iii) Árvore de *frames*
A estrutura de dados que mais se aproxima da rede de *frames* é a árvore, como já foi discutido na seção 4.4, porém nem todas as propriedades desta estrutura são respeitadas na FrameNet Brasil: filhos com mais de um pai, irmãos com ligações diretas (sem passar pelo pai), ciclos entre nós de uma subárvore dentre outros, são os casos que podem ser verificados nessa rede. O especialista afirmou que prender a FrameNet Brasil a uma estrutura tão rígida limita o projeto no que tange a representação dos mais variados cenários da vida cotidiana, bem como na associação de informações que agreguem valor semântico ao *frame*. No entanto, quando foram apresentadas as subárvores com restrições violadas, foram necessárias algumas análises semânticas para validar os casos. A Figura 6(b) ilustra um exemplo de *frames* irmãos se relacionando sem passar, necessariamente, pelo *frame* pai: os *frames* `inibir_o_movimento` e `liberar` *herdam* do *frame* `afetar_intencionalmente`, porém eles se ligam pela relação de *precedência* que define a ordem em que os fatos ocorrem.

## 5. Considerações Finais

Este trabalho apresentou uma análise da rede de *frames* da FrameNet Brasil, com o objetivo de identificar situações que pudessem estar em desacordo com o modelo

conceitual da mesma.

Foram aplicadas medidas de centralidade e algoritmos de agrupamento na análise da rede. Não foram encontrados casos explícitos de violação do modelo conceitual da FrameNet Brasil porém, foram identificados cenários atípicos que foram repassados aos profissionais do projeto para uma análise detalhada.

Como trabalhos futuros pretende-se avançar na análise da rede de *frames* usando métricas mais sofisticadas de redes complexas e usar a FrameNet Brasil como uma ferramenta de Análise de Sentimentos.

## Referências


[Baker et al. 1998] Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.

[Barabási et al. 2000] Barabási, A.-L., Albert, R., and Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications*, 281(1-4):69–77.

[Blondel et al. 2008] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

[Fillmore and Baker 2010] Fillmore, C. J. and Baker, C. (2010). A frames approach to semantic analysis. In *The Oxford handbook of linguistic analysis*.

[Fillmore et al. 2004] Fillmore, C. J., Baker, C. F., and Sato, H. (2004). Framenet as a"net". In *LREC*.

[Fillmore et al. 2003a] Fillmore, C. J., Johnson, C. R., and Petruck, M. R. (2003a). Background to framenet. *International journal of lexicography*, 16(3):235–250.

[Fillmore et al. 2003b] Fillmore, C. J., Petruck, M. R., Ruppenhofer, J., and Wright, A. (2003b). Framenet in action: The case of attaching. *International journal of lexicography*, 16(3):297–332.

[Ruppenhofer et al. 2016] Ruppenhofer, J., Ellsworth, M., Petruck, M. R., Johnson, C. R., and Scheffczyk, J. (2016). *FrameNet II: Extended theory and practice*. Institut für Deutsche Sprache, Bibliothek.

[Salomao 2009] Salomao, M. M. M. (2009). Framenet brasil: um trabalho em progresso. *Calidoscópio*, 7(3):171–182.

[Torrent et al. 2014] Torrent, T., Salomão, M. M., Campos, F., Braga, R., Matos, E., Gamonal, M., Gonçalves, J., Souza, B., Gomes, D., and Peron, S. (2014). Copa 2014 framenet brasil: a frame-based trilingual electronic dictionary for the football world cup. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*, pages 10–14.

[Wasserman and Faust 1994] Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.