Relatório do diagnóstico da ontologia atualmente existente no Ministério da Cultura voltada para informações e indicadores culturais

Dalton Lopes Martins

Marcel Ferrante Silva

Luís Felipe Rosa de Oliveira

1. Introdução

O Ministério da Cultura (MinC) foi criado no ano de 1985 como fruto de um conjunto diverso de ações oriundas do processo de abertura democrática pelo qual passava o Brasil de então. Tem-se, portanto, uma história organizacional de mais de 30 anos composta por diferentes gestões, visões de política pública para a Cultura, processos de trabalho, pessoas e, como consequência de sua própria dinamicidade social, um conjunto expressivo de sistemas e estratégias de organização da informação.

Não se pode falar de um diagnóstico de processos informacionais sem levarmos em consideração essa história social a partir da qual esses processos são reflexo. É somente a partir do reconhecimento dessa dinâmica que a informação e seus modos de organização ficam mais visíveis e podemos com eles analisar a própria instituição da qual são produtos.

Consideramos que uma forma de olhar para essa história e que será considerada como ponto de partida desse diagnóstico que aqui se apresenta, é observarmos quais são as fontes e as formas de coleta de informação institucional hoje existentes e utilizadas no MinC. Essas fontes foram produtos de inúmeras decisões organizacionais e refletem a necessidade de diferentes políticas, sendo, portanto, elementos fundamentais de operacionalização dessas políticas e por onde elas se estruturam como pontos de diálogo com a sociedade, processos de implementação, seleção, acompanhamento e avaliação.

Dessa maneira, o que apresentamos como diagnóstico neste relatório é uma análise das fontes de informação coletadas pela Coordenação Geral de Monitoramento de Informações Culturais da Secretaria de Políticas Culturais e que foram consideradas por essa coordenação como o produto de diferentes estratégias já realizadas pelo MinC na direção da construção de taxonomias e vocabulários controlados.

O relatório está dividido nas seções apresentadas a seguir. **Objetivos específicos**, ressalta quais são os principais resultados e estratégias utilizadas para a realização deste relatório. **Revisão de documentos de referência**, apresenta cada fonte de informação considerada para nossa análise, bem como quais as referências

conceituais e técnicas que utilizamos para as recomendações realizadas ao final do relatório. **Análise de dados**, apresenta a análise do total de expressões utilizadas para estruturar cada fonte de informação, dividindo essas expressões já em candidatos a termos para compor a ontologia e candidatos a atributos dos termos. Ao longo dessa seção, explicamos em detalhes como essa divisão foi realizada e suas possíveis consequências para o trabalho colaborativo de desenvolvimento da ontologia. Também nesta seção, apresentamos uma análise da rede semântica formada pelas expressões consideradas candidatos a termos. Essa estratégia visa facilitar a visualização e mapeamento dos pontos de contato entre as diferentes fontes de informação, já facilitando a integração entre essas fontes no trabalho de concepção da ontologia. **Diagnóstico dos problemas técnicos,** apresentamos um os problemas técnicos identificados nas expressões candidatas a termos nas fontes de informação atual com base nas normas técnicas utilizadas. Por fim, **Considerações**, propomos algumas conclusões que esse diagnóstico nos permitiu chegar, bem como sugestões de encaminhamentos que deveriam ser consideradas pelo MinC para as etapas futuras do processo de construção da ontologia.

2. Objetivos específicos

a. Identificar termos e atributos dos termos candidatos para a composição da ontologia de políticas culturais;
b. Mapear quais são os candidatos a termos que possuem maior conectividade entre as diferentes fontes de informação e podem servir de elementos transversais para composição da ontologia, ampliando seu potencial de coesão estrutural;
c. Mapear os buracos estruturais entre as fontes de informação na análise de sua rede semântica dos termos candidatos de forma a facilitar a identificação de pontos de convergência que devem ser levados em consideração na composição da ontologia;
d. Identificar os problemas técnicos das expressões atualmente em uso segundo orientações de normas internacionais para a construção de termos de taxonomias e produzir recomendações que evitem a recorrência desses problemas em trabalhos futuros;
e. Produzir recomendações técnicas que favoreçam as estratégias de mediação do trabalho colaborativo de desenvolvimento da ontologia.

3. Revisão dos documentos de referência

Utilizamos dois tipos de documentos de referência para a construção deste diagnóstico: fontes de informação e normas técnicas/trabalhos científicos.

As **fontes de informação** são as bases de dados[1], formulários[2] e pesquisas de referência coletadas pela Coordenação Geral de Monitoramento de Informações Culturais e repassado a equipe da Universidade Federal de Goiás como sendo as referências por meio das quais as várias iniciativas de construção de vocabulários controlados, glossários e taxonomias hoje se expressam. Vale mencionar que essa coleta se deu por solicitação da coordenação às diferentes áreas do MinC, demandando todos os formulários e bases de dados que essas áreas utilizam como forma de organização de suas informações estratégicas. Partimos do princípio, portanto, que as fontes consideradas em nosso diagnóstico passaram por um crivo das áreas organizacionais do MinC, logo filtrando e selecionando aquilo que consideraram de maior relevância para sua atuação institucional.

As **normas técnicas/trabalhos científicos** representam as referências técnicas de pesquisadores, grupos de pesquisas e instituições especializadas na produção de normas construídas em amplos processos de debate e consulta a especialistas na área de Ciência da Informação como maneira de indicar as melhores práticas para a construção de termos para taxonomias. Observar as expressões semânticas das fontes de informação por meio dessas normas produz uma avaliação crítica do trabalho produzido na composição dessas fontes e ao mesmo tempo facilita a identificação de práticas que devem ser consideradas para as etapas futuras de desenvolvimento da ontologia de políticas culturais.

a. Fontes de informação

A Coordenação Geral de Monitoramento de Informações Culturais coletou as seguintes fontes de informação das áreas do MinC:

1. **Do Instituto Nacional de Museus (IBRAM):**
    a. Pesquisa "Conheça os museus brasileiros": é uma pesquisa oriunda do Instituto Brasileiro de Museus (IBRAM) e realizada a partir de um formulário de mapeamento e atualização de informações de cada museu realizada pelos próprios museus e encaminhada ao IBRAM. O formulário pode ser encontrado neste link: http://www.museus.gov.br/wp-content/uploads/2015/09/Formulario-Mapeamento-e-Atualizacao-de-Informacoes.docx
    b. Cadastro nacional de museus: são os dados coletados pelo IBRAM sobre os museus brasileiros e disponibilizados neste link: http://www.museus.gov.br/sistemas/cadastro-nacional-de-museus/ .

[1] Bases de dados no contexto deste relatório podem ser planilhas de cálculo utilizadas por diferentes áreas do MinC para organização de informações estratégicas ou bases de dados estruturadas em linguagem SQL e utilizadas por diferentes sistemas de informação do MinC.

[2] Formulários podem ser fichas de cadastro preenchidas por técnicos do MinC para avaliação ou diagnóstico de alguma etapa de uma política, ou então por usuários das políticas públicas do MinC quando da inscrição em editais, concursos ou prêmios.

c. Dicionário de dados: é a estrutura de dados utilizada pelo sistema de informação do Cadastro Nacional de Museus. Pode ser acessado neste link: http://sistemas.museus.gov.br/cnm/pesquisa/filtrarUf.
d. Planilha para recarga de dados CNM – SNIIC: é a planilha utilizada para compatibilizar o sistema do Cadastro Nacional de Museus e o antigo Sistema Nacional de Informações e Indicadores Culturais (SNIIC). Essa planilha fazia a conversação de campos entre um sistema e outro.
e. Bens desaparecidos: são dados relativos a caracterização do que são bens desaparecidos para objetos musealizados.
f. Tipologia de Eventos: tipologia de eventos utilizadas para ações realizadas pelos Museus.

2. **Da Diretoria Livro, Leitura, Literatura e Bibliotecas (DLLLB):**
   a. Ficha de inscrição Viva Leitura: é a ficha de inscrição do Prêmio Viva Leitura. Pode ser acessada aqui: http://www.premiovivaleitura.org.br/cadastro/cadastro.ASP.
   b. Pesquisa Leitura para Todos: pesquisa relacionada ao Prêmio Leitura para Todos da Biblioteca Nacional.
   c. Cadastro formulário de bibliotecas públicas: é a ficha utilizada pelo Sistema Nacional de Bibliotecas para mapeamento das bibliotecas públicas brasileiras. Os dados estão estruturados neste link: http://snbp.culturadigital.br/informacao/dados-das-bibliotecas-publicas/ .
3. **Dos Equipamentos Culturais:**
   a. Base de estrutura de dados da SAV: é a base de dados utilizada para organizar as informações da Secretaria do Áudio Visual do MinC.
   b. Formulário da rede Cultura Viva: é o formulário utilizado pelos Pontos de Cultura para integração a Rede Cultura Viva.
   c. Taxonomia SNIIC – Centro de Artes e Esportivos Unificados (CEUs): é a taxonomia utilizada pelo SNIIC para definição de tipos de CEUs.
   d. Teatros do Brasil (SNIIC): é a taxonomia utilizada pelo SNIIC para definição dos tipos de teatros brasileiros.
4. **Do Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN):** todos os itens abaixo foram enviados pelo IPHAN como fichas cadastrais que expressam os dados necessários para tipos específicos de patrimônios que fazem parte do acervo do instituto.
   a. Ficha do bem imóvel.
   b. Ficha do bem paisagístico.
   c. Ficha dos bens móveis e integrados.
   d. Ficha do conjunto arquitetônico.
   e. Ficha do conjunto urbano.
   f. Ficha de edificação.
   g. Ficha do patrimônio ferroviário.

5. **Da Agência Nacional de Cinema (ANCINE):**
    a. Tipologia da ANCINE: tipologia de utilizada para categorizar complexos de exibição de filmes, agentes econômicos, segmentos do mercado audiovisual e nomenclaturas e definições da agência.
6. **Do Mapas Culturais:** as fontes de informação do mapas culturais são relativas as diferentes formas de categorização e tipologias dos 4 elementos estruturantes que são cartografados pelo mapas: agentes, espaços, projetos, eventos e esferas. O Mapas Culturais pode ser acessado aqui: http://mapas.cultura.gov.br/.
    a. Agentes.
    b. Espaços.
    c. Projetos.
    d. Eventos.
    e. Esfera.
7. **Do Sistema de apoio às Leis de Incentivo à Cultura (SALIC):**
    a. Produtos e áreas culturais: é toda a base de dados do SALIC com suas tipologias e vocabulários controlados para as áreas e produtos culturais. O SALIC pode ser acessado aqui: http://novosalic.cultura.gov.br/.
    b. Equipamentos culturais: é a base de equipamentos culturais do SALIC.

Vale observar que a forma de organização das fontes de informação já revela uma dificuldade inicial de localizar a origem específica de cada fonte, entendo que muitas delas servem a diferentes projetos e estão incorporadas em sistemas de informação distintos e transversais às áreas institucionais do MinC. Já se pode observar que várias redundâncias na representação de informação, como o caso dos equipamentos culturais que aparecem mais detalhadamente representados no item 3 acima por diferentes cadastros, fichas e bases de áreas específicas do MinC, como a SAV, bibliotecas públicas e museus, e também vemos equipamentos culturais aparecem novamente no item 7, dessa vez ligados ao sistema SALIC. Outra redundância que podemos observar é da tipologia de eventos no item 01 relacionados aos Museus e dos eventos no item 6 como elementos dos Mapas Culturais.

Entendemos que é exatamente a busca por padronizar essas redundâncias, encontrando seus pontos de convergência e, em casos específicos, de divergência no que consiste o trabalho de desenvolvimento de uma ontologia de políticas culturais para o MinC. Ao percebermos essas redundâncias, fica ao mesmo tempo visível que se temos diferentes fontes de informação representando o mesmo objeto informacional e, o que avaliaremos apenas na próxima sessão, eventualmente representando esse objeto de forma diferente, teremos enorme dificuldade de integrar essas informações no mesmo espaço. A falta de integração dificulta ter uma visão integrada das diferentes políticas culturais, impedindo muitas estratégias de planejamento estratégico, tais como

avaliação de impacto de políticas, que poderiam potencializar as próprias ações do MinC.

b. Normas técnicas e trabalhos científicos

Para a realização deste trabalho de diagnóstico, utilizamos 3 fontes de informação técnicas e científicas:

- A norma técnica ANSI/NISO Z39.19-2005 Revisão de 2010 – *Guidelines for the construction, format and management of monolingual controlled vocabularies* – Disponível para acesso em: http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf. Essa norma foi construída com base em várias normas técnicas norte-americanas e internacionais numa tentativa de integrar e padronizar a forma de construção de vocabulários controlados. Entendemos ser uma referência para a realização deste trabalho.
- O livro "*Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*" de autoria de Patrick Lambe, editora Chandos Publishing, de 2014. O livro apresenta uma das poucas revisões científicas de construção de taxonomias que dialoga com processos institucionais de trabalho, sendo uma referência para a crítica da relação entre estruturas de representação e organização da informação e modos institucionais de gestão e trabalho.
- O livro "*The Accidental Taxonomist*" de autoria de Heather Hedden, editora Information Today, de 2010. O livro apresenta uma abordagem dos processos de construção de taxonomias e ontologias de maneira simples e orientada a público leigo, colocando ênfase nos diferentes profissionais do mundo das organizações que possivelmente se envolvem em trabalhos de construção de estruturas de representação da informação. O texto facilita identificar alguns cuidados que são necessários tomar quando da construção colaborativa de uma ontologia, bem como fornece preciosas indicações de modos de análise de taxonomias.

4. Análise dos dados

O trabalho de análise de dados das fontes informacionais apresentadas anteriormente teve por objetivo identificar todas as expressões utilizadas nos documentos fornecidos e identificar aquilo que poderia ser considerado como um possível termo para compor a ontologia e o que poderia ser atributos de termos, entendendo por atributo elementos descritivos de dados que ajudam a caracterizar o que é um determinado termo da ontologia.

A metodologia de análise utiliza é relatada nos passos a seguir:

- Coleta dos documentos que compõem cada fonte de informação;

- Análise dos documentos para identificação de expressões;
- Registro das expressões em uma planilha de dados organizada por fonte de informação, documento e nível hierárquico de origem da informação. Vale ressaltar que o nível hierárquico é relativo ao modo de apresentação da informação no documento de origem, sendo, por exemplo, o estado de Goiás subordinado ao termo Estados em um formulário de dados;
- Classificação das expressões por termos e atributos candidatos na planilha de dados. Os termos são aquelas expressões semânticas que compõem diferentes tipologias para definir objetos de informação e os atributos são expressões que definem dados, qualitativos e quantitativos, específicos para determinadas expressões. Um exemplo de termo seria o nome de um povo indígena para caracterizar um tipo de população e um exemplo de um atributo seria o número do CPF de um participante de um edital;
- Construção de tabelas comparativas por fontes de informação e por documentos;
- Construção das redes semânticas por documentos e integrada por toda as fontes de informação;
- Identificação dos nós centrais da rede semântica integrada como elementos que sugerem termos transversais a todo o espaço informacional e que podem apoiar a construção da ontologia;
- Identificação de buracos estruturais como elementos que sugerem pontos de atenção para conectar áreas da ontologia ainda desconectadas.

Apresentamos o resultado da análise de dados nos itens a seguir.

a. Tabelas de expressões

O detalhamento das tabelas por se tratar de uma grande quantidade de dados é apresentado de forma digital e pode ser acessado pelo link: https://drive.google.com/file/d/0ByyRLUo5xMDKQ3ZoM1JJT2FHUjQ/view?usp=sharing ou pelo microlink http://tinyurl.com/gl6u8vt. Neste link são apresentadas 08 abas de informação, sendo as 07 primeiras onde constam as expressões analisadas e a última uma síntese descritiva das informações, apresentada nos itens a seguir.

b. Tabelas comparativas

Uma vez tendo identificado os possíveis termos e atributos candidatos a partir da análise das tabelas apresentadas no item anterior, realizamos uma comparação em duas etapas, sendo a primeira entre as áreas do MinC apresentado na tabela 01 e a segunda entre as áreas e com maior detalhamento por fontes de informação coletada apresentada na tabela 02. As tabelas o total de expressões identificadas, o total de termos candidatos selecionados pela equipe da pesquisa, o porcentual que esses termos representam em relação ao

total de expressões, o total de atributos selecionados pela equipe de pesquisa e o porcentual que esses atributos representam em relação ao total de expressões.

É importante ressaltar que o que consideramos termos candidatos nessa análise são termos que sugerimos que devem ser analisados pelos processos de curadoria da construção da ontologia, podendo ou não serem aproveitados em etapas futuras. No entanto, a análise e o debate sobre sua importância são estratégicos para valorizar a própria memória institucional do que consiste o atual estágio de estruturação informacional do MinC.

Base	Fontes de informação	Total de Expressões	Candidatos a Termos	%	Candidatos a Atributos	%
Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN)	7	1286	621	48%	665	52%
Mapas Culturais	5	788	758	96%	30	4%
Equipamentos Culturais	4	724	401	55%	323	45%
Instituto Nacional de Museus (IBRAM)	6	720	411	57%	309	43%
Sistema de apoio às Leis de Incentivo à Cultura (SALIC)	2	615	295	48%	320	52%
Diretoria Livro, Leitura, Literatura e Bibliotecas (DLLLB)	3	356	121	34%	235	66%
Agência Nacional de Cinema (ANCINE)	1	47	47	100%	0	0%
TOTAL GERAL		**4536**	**2654**	**59%**	**1882**	**41%**

Tabela 01 – Candidatos a termos e atributos por macro fontes de informação. Fonte: Dados da pesquisa.

Vemos na tabela 01 que as origens institucionais em relação ao MinC estão ordenadas na coluna Base por ordem do maior para o menor em termos do total de expressões identificadas. Temos aqui um total de 4536 expressões identificadas. Vemos que o IPHAN é a maior fonte de informação e a ANCINE a menor, apresentando uma variação expressiva de termos entre ambos. No entanto, os intermediários possuem uma quantidade de expressões bastante próxima, girando em torno de 700 expressões.

A coluna “Fontes de Informação” mostra a quantidade de documentos, sejam eles bases de dados, formulários ou pesquisas, fornecidas por área do MinC representada na coluna “Base”. Vemos novamente o IPHAN com a maior quantidade de fontes, seguido pelo IBRAM e o Mapas Culturais.

Já quando analisamos a distribuição das expressões entre termos e atributos, temos uma redução geral expressiva de termos para 59% do número original de expressões, chegando a um total de 2654 termos candidatos para a composição da ontologia. Vale dizer que essa redução é extremamente importante, pois reduz e facilita o foco de atenção no trabalho de curadoria desses termos, otimizando os processos de discussão e avaliação daquilo que será prioritário e estruturante para a composição da ontologia de políticas culturais.

A maior redução de termos se deu no âmbito da DLLLB, seguido por IPHAN e SALIC e a menor na ANCINE, onde todos as expressões foram consideradas termos candidatos.

Na tabela 02, a seguir, apresentamos os dados detalhados por fonte de informação.

Base	Documento	Total de Expressões	Candidatos a Termos	%	Candidatos a Atributos	%
Instituto Nacional de Museus (IBRAM)	Conheça os museus brasileiros	71	41	58%	30	42%
	Cadastro Nacional de Museus (CNM)	104	39	38%	65	63%
	Dicionário de dados	26	26	100%	0	0%
	Planilha para recarga de dados CNM - SNIIC	124	46	37%	78	63%
	Bens Desaparecidos	303	215	71%	88	29%
	Eventos	92	44	48%	48	52%
	TOTAL	**720**	**411**	**57%**	**309**	**43%**
Diretoria Livro, Leitura, Literatura e Bibliotecas (DLLLB)	Ficha de inscrição viva leitura	54	5	9%	49	91%
	Cadastro Formulário Biblioteca Pública (BP)	257	92	36%	165	64%
	Pesquisa leitura para todos	45	24	53%	21	47%
	TOTAL	**356**	**121**	**34%**	**235**	**66%**
Equipamentos Culturais	Base de Estrutura de Dados SAV v2	65	14	22%	51	78%
	Formulário da Rede Cultura Viva	270	189	70%	81	30%
	Taxonomia SNIIC – Centro de Artes e Esportes Unificados (CEUs)	194	186	96%	8	4%
	Teatros Brasil – SNIIC	195	12	6%	183	94%
	TOTAL	**724**	**401**	**55%**	**323**	**45%**
Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN)	Bem Imóvel Ficha	249	125	50%	124	50%
	Bem Paisagístico Ficha	311	116	37%	195	63%
	Bens Móveis e Integrados	70	27	39%	43	61%
	Conjunto Arquitetônico Ficha	180	80	44%	100	56%
	Conjunto Urbano Ficha	153	105	69%	48	31%
	Edificação Ficha	259	132	51%	127	49%
	Patrimônio Ferroviário	64	36	56%	28	44%
	TOTAL	**1286**	**621**	**48%**	**665**	**52%**
Agência Nacional de Cinema (ANCINE)	Tipologia SNIIC MINC Agência Nacional do Cinema (ANCINE)	47	47	100%	0	0%
	TOTAL	**47**	**47**	**100%**	**0**	**0**
Mapas Culturais	Agentes	407	407	100%	0	0%
	Espaços	158	136	86%	22	14%
	Projetos	42	34	81%	8	19%
	Eventos	69	69	100%	0	0%
	Esfera	18	18	100%	0	0%
	TOTAL	**788**	**758**	**96%**	**30**	**4%**
Sistema de apoio às Leis de Incentivo à Cultura (SALIC)	Produtos e áreas culturais	155	155	100%	0	0%
	Equipamentos Culturais	460	140	30%	320	70%
	TOTAL	**615**	**295**	**48%**	**320**	**52%**
TOTAL GERAL		**4536**	**2654**	**59%**	**1882**	**41%**

Tabela 02 – Candidatos a termos e atributos por fontes de informação. Fonte: Dados da pesquisa.

Vemos na tabela 02 o detalhamento dos documentos que compõem cada fonte de informação. A maior fonte de informação é a tipologia de “Equipamentos Culturais” do SALIC com 460 expressões sendo 140 consideradas termos candidatos e a segunda maior é a tabela de “Agentes” que compõem o Mapas Culturais com 407 expressões, sendo todas elas consideradas como termos candidatos. A menor fonte de informação é também a tabela “Esfera” do Mapas Culturais com 18 expressões sendo todas consideradas termos candidatos seguida pelo “Dicionário de Dados” do IBRAM com 26 expressões sendo todas também consideradas como termos candidatos.

Entendemos que temos aqui uma enorme riqueza semântica que representa décadas de construção de estruturas informacionais por parte do MinC. Essa é uma forma de dar visibilidade a essa variabilidade semântica a memória institucional de uma organização, procurando ressaltar os termos e atributos que foram usados ao longo dos anos para representar suas informações e, sem dúvida, organizar os diferentes modos de pensar e conceber a política pública na área cultural.

De maneira a facilitar entendermos melhor como se manifesta esses modos de organização, apresentamos na seção a seguir as redes semânticas, que mostram como os termos se relacionam entre si por fonte de informação, conforme apresentado na tabela 02. Ao final, também apresentamos uma rede que interconecta todas as fontes de informação, de onde teremos uma visão geral do espaço informacional semântico do MinC atualmente em uso.

c. Análise das redes semânticas

As redes semânticas são uma forma de representarmos visualmente as relações entre os termos candidatos de cada fonte de informação, facilitando identificarmos como esses termos estão estruturados. A rede é construída representando visualmente cada termo como um nó e a relação entre os termos como uma conexão. Ao compararmos as várias fontes de informação dessa maneira podemos notar as diferenças e semelhanças entre elas, bem como suas estratégias de organização.

A seguir, apresentamos uma rede semântica que integra todas as fontes de informação. A maior vantagem dessa integração é encontrarmos os termos que estão sobrepostos em cada fonte de informação, ou seja, termos que aparecem em mais de uma fonte e são, portanto, conectores naturais entre as diferentes fontes. É importante ressaltar que ao encontrarmos essas conexões estamos encontrando convergência semântica entre elas, o que deve ser entendido como a identificação da mesma maneira de nomear objetos. Sem dúvida, não é possível apenas com essa análise sabermos se ao usar o mesmo nome o significado atribuído a esse nome é o mesmo. Porém, já temos aqui ao identificar essas convergências importantes elementos de conversação pelo simples fato de termos a oportunidade de compartilhar os significados que cada área atribui a esses nomes. É nesse espaço de debate que as estratégias de mediação do trabalho de formulação da ontologia vão ganhando expressão e sentido para seus participantes. Apresentamos no apêndice 1 cada rede semântica por fonte de informação.

A figura 01 apresenta a rede semântica integrada a partir de todas as fontes de informação apresentadas na tabela 02. A rede possui 2167 nós e 3020 conexões entre eles. É importante ressaltar que tínhamos inicialmente na tabela 02 um número de 2654 termos candidatos e agora na rede 2167 nós, sendo que os nós representam termos candidatos. O que faz com que ocorra essa redução 487 termos é o fato de que muitos termos aparecem mais de uma vez pelas diferentes de informação e quando representados em rede esses termos redundantes desaparecem e aparecem uma única vez, porém representando as diferentes conexões desses termos e suas fontes de origem. Esses são termos extremamente estratégicos pois são eles que formam as pontes que conectam as fontes de informação e já apresentam um comportamento transversal no espaço informacional, logo devem ser elementos prioritários de serem considerados como termos candidatos para a constituição da ontologia. Logo mais,

detalharemos esses termos e apresentamos uma forma de priorizá-los e entender seu papel estrutural na rede semântica.

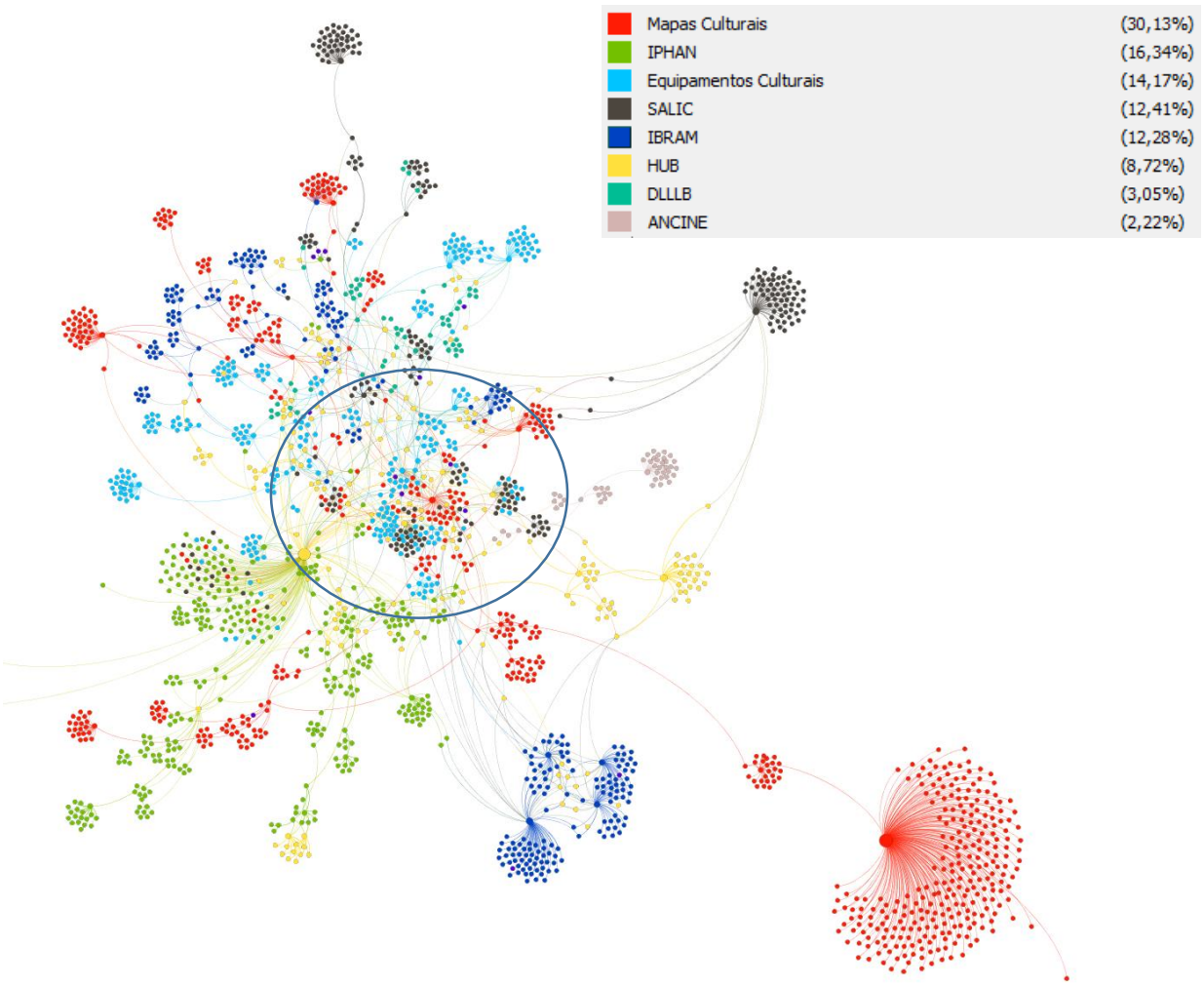


Figura 01 – Rede semântica integrada das diferentes fontes de informação. Fonte: Dados da pesquisa.

Observando a figura 01, temos que as cores representam as áreas institucionais do MinC e são o elemento mais importante de organização dos nós da rede semântica. Temos 8 cores no grafo e a legenda no canto superior da figura mostra o que significa cada cor e quantidade relativa de nós presentes no grafo respectivo a essa cor. Temos que o “Mapas Culturais” é a fonte de informação mais expressiva na constituição da rede semântica, chegando a um número de mais de 30% dos seus nós, sendo seguida por IPHAN e “Equipamentos Culturais”. Os termos candidatos que definem cada nó da rede podem ser vistos em maiores detalhes por fonte de informação no apêndice 01.

Destacamos de forma circular o centro do grafo, entendo que essa parte da rede semântica é uma das mais importantes, pois é por meio dela que os demais elementos periféricos possuem conexões entre si. O centro de um grafo é considerado sua parte estrutural que garante o padrão de conectividade da rede como um todo, sendo ele o ponto de convergência e conexão necessária quando do fluxo de informação de uma

ponta da rede a outra. O que notamos de maneira evidente nessa parte destacada é um predomínio das cores azul em maior volume, seguido pelo vermelho e cinza escuro, em menor quantidade. São respectivamente as fontes de informação “Equipamentos Culturais”, “Mapas Culturais” e “SALIC”. Vamos nos ater mais detalhadamente nesse resultado. O que se pode afirmar a partir dessa constatação é que o centro da rede semântica constituída pelas principais fontes de informação do MinC é dominada majoritariamente pelos sistemas de informação “Mapas Culturais” e “SALIC” e os termos candidatos dos “Equipamentos Culturais” possuem expressiva presença entre eles.

Não é escopo deste relatório investigar as causas dessa presença, porém cabe notar que essa forma de ler o espaço informacional levanta questões sobre a própria forma de constituição da política cultural do MinC nas últimas décadas. Seriam de fato os equipamentos culturais e suas formas organizacionais um dos elementos mais importantes da política cultural do MinC e seria isso que justificaria essa presença expressiva como elementos estruturantes do centro da rede semântica desse ministério? Que questões políticas são dadas estão ligadas a esses equipamentos para ter essa relevância informacional? E os “Mapas Culturais” e “SALIC”, são eles também sistemas vitais de organização do espaço informacional do MinC e por isso compõem o centro dessa rede semântica? O que significa esse espaço ocupado por essas 3 fontes de informação? Há alguma sobreposição ou composição política entre eles? Enfim, podemos aqui levantar inúmeras questões que valem deixar registrado nesse relatório, pois futuras pesquisas sobre o histórico das práticas informacionais do MinC podem se valer das constatações que aqui apresentamos.

Para efeito do trabalho de diagnóstico das atuais estruturas de informação do MinC e mesmo para os trabalhos futuros de constituição da ontologia, cabe afirmar aqui que o centro dessa rede semântica são os elementos estruturais mais importantes das estratégias de informação atualmente vigente no MinC. Logo, deveriam ser elas priorizadas como focos de atenção para análise de seus termos candidatos e de seus atributos para a ontologia de políticas culturais. Dialogar o mesmo dialeto desse centro da rede significa, de partida, se posicionar de forma estratégica na mesma posição, logo, podendo dialogar de forma central com as outras fontes de informação.

De maneira a ressaltar ainda mais a importância desse centro da rede, apresentamos nas figuras 02 e 03 a mesma rede semântica, sendo na primeira sem a presença dos nós oriundos do “Mapas Culturais” e na segunda sem a presença dos nós dos “Equipamentos Culturais”. Vejamos.

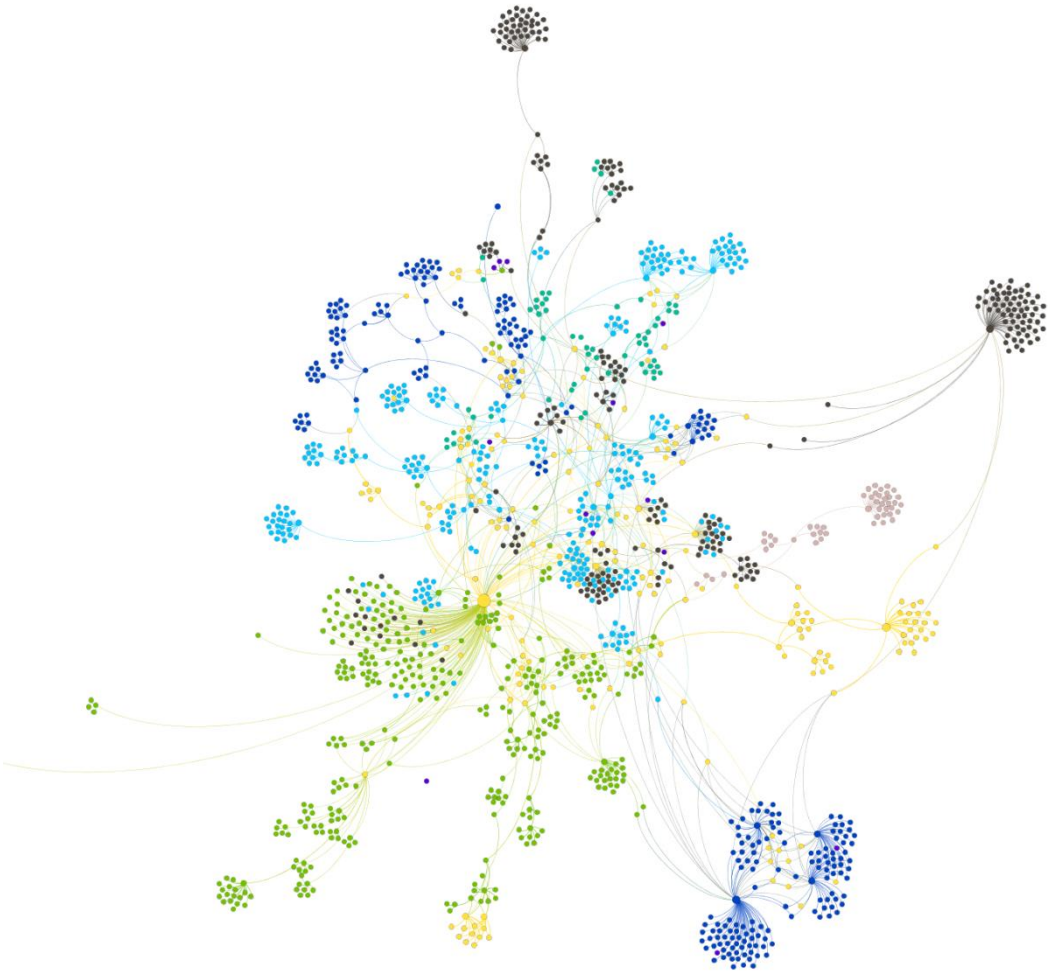


Figura 02 – Rede semântica integrada das diferentes fontes de informação sem a fonte “Mapas Culturais”. Fonte: Dados da pesquisa.

Analisando a figura 02 com a figura 01, vemos como a rede não apenas perde em torno de 30% dos seus nós, mas como o centro se torna menos denso e concentrado apenas em torno do “SALIC” e “Equipamentos Culturais”.

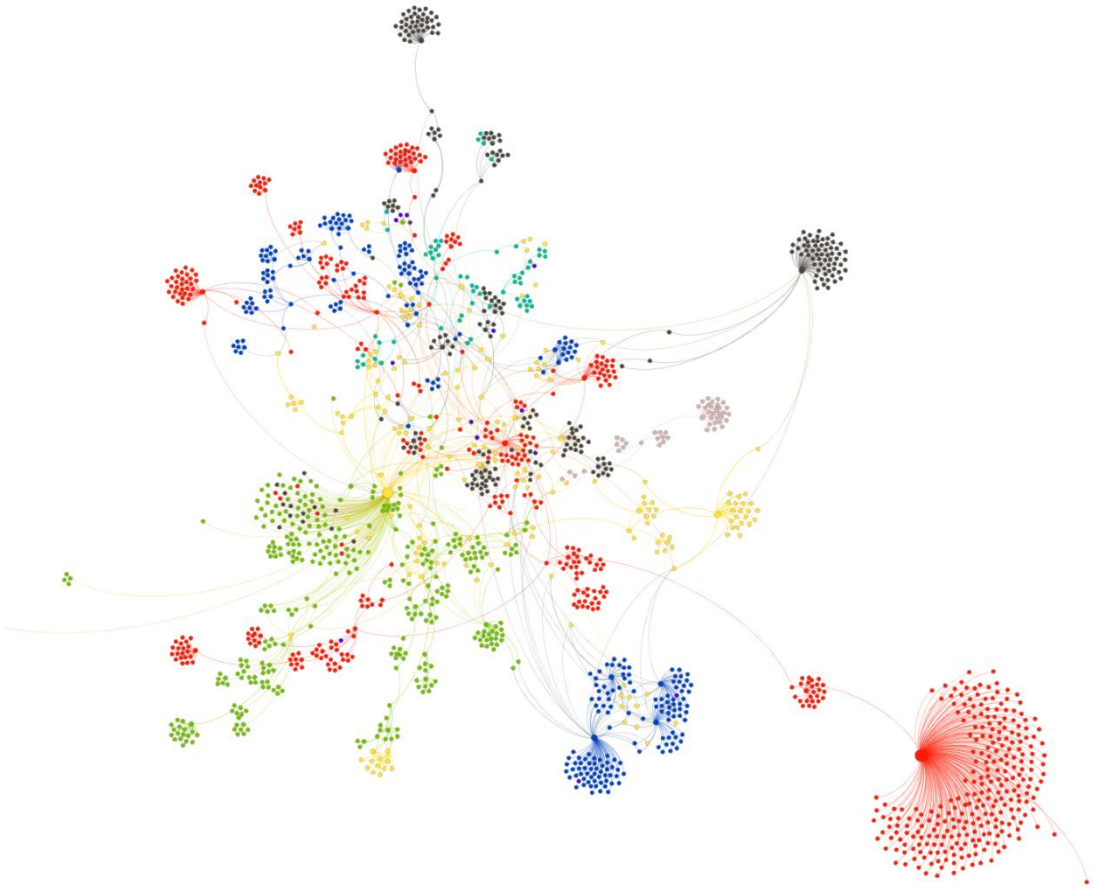


Figura 03 – Rede semântica integrada das diferentes fontes de informação sem a fonte “Equipamentos Culturais”. Fonte: Dados da pesquisa.

Sem a presença da fonte “Equipamentos Culturais”, o centro da rede semântica passa a ser dominado quase que exclusivamente pelos “Mapas Culturais” com alguma presença do “SALIC”. O impacto no centro da rede da saída dos equipamentos é maior do que da saída dos mapas, o que confirma as impressões anteriores onde apontamos que os equipamentos culturais parecem ser o conjunto de termos candidatos mais importantes da estruturação informacional do MinC.

Utilizando como indicador de centralidade dos termos candidatos, utilizamos o algoritmo de cálculo de centralidade por intermediação para calcular o valor de cada termo candidato na rede semântica. Esse cálculo aponta quais são os termos candidatos mais intermediários entre todos, ou seja, aqueles que são mais centrais nos caminhos de conectividade relativo entre todos os termos. Apresentamos na tabela 03 os 20 termos mais importantes, bem como de que fonte de informação são oriundos e o valor de sua centralidade.

Termo candidato	Fonte de Informação	Centralidade
Sistema de gerenciamento eletrônico do acervo	SALIC	9.724079251245898E-5
Período de Construção	IPHAN	9.596130840045293E-6
Recursos Humanos	Equipamentos Culturais	9.382883488044287E-6
Identificação numérica do Tipo da Esfera Administrativa	Equipamentos Culturais	9.382883488044287E-6
Complexos	SALIC	9.340234017644086E-5
Abastecimento de água	IPHAN	9.265597444443733E-5
Biblioteca	HUB	9.153997996896541E-4
Agentes	Mapas Culturais	9.002592376975826E-5
Cobertura	HUB	8.992742380240542E-5
Equipamentos Culturais e Meios de Comunicação	Equipamentos Culturais	8.956388784042275E-6
Caracterização Interna	IPHAN	8.743141432041267E-6
Dados sobre Serviços e Atividades	SALIC	8.606663126760622E-5
Tipo de Pavimentação	IPHAN	8.529894080040261E-7
Categoria de Trabalho	DLLLB	8.529894080040261E-7
Colaboradores	Equipamentos Culturais	8.529894080040261E-7
Sistema de Classificação	DLLLB	8.529894080040261E-7
Serviços de Informação e Atendimento	DLLLB	8.529894080040261E-7
Participação em eventos municipais na área da cultura	DLLLB	8.529894080040261E-7
Formulário de Biblioteca Comunitária/ Ponto de Leitura	DLLLB	8.529894080040261E-7
Patrocinadores Financiadores e Incentivadores Culturais	Mapas Culturais	8.529894080040261E-6

Tabela 03 – 20 termos candidatos de maior centralidade por intermediação na rede semântica. Fonte: Dados da pesquisa.

Vemos na tabela 03 os 20 termos mais centrais por intermediação da rede semântica. Vale dizer que constam nessa tabela não apenas os termos do centro da rede, pois alguns podem ter algo valor de centralidade em um pedaço periférico da rede, pois podem ser termos que conectam muitos outros que dele dependem como caminho de conexão com outros pedaços da rede. Esses termos devem ser considerados como elementos estratégicos e também prioritários para a produção da ontologia.

5. Diagnóstico dos problemas técnicos

A análise dos problemas técnicos leva em consideração as normas técnicas ANSI/NISO Z39.19. Separamos os problemas e sugestões de melhorias nos tópicos a seguir:

a. **Caixa alta e baixa:** a norma técnica recomenda que as expressões sejam todas descritas em caixa baixa e a utilização de caixa alta apenas na primeira letra dos termos. Há vários termos utilizados pelas fontes de informação que misturam caixa alta ao longo de várias palavras, como por exemplo, “Serviços de Informação e Atendimento”, na tabela 03 acima, entre outras expressões. Recomenda-se a normalização dessa forma de escrita;
b. **Termos simples ou de palavras múltiplas:** recomenda-se que termos tenham no máximo de uma a quatro palavras. Temos vários exemplos pelas fontes de informação que acabam não seguindo essa regra, como “Formulário de Biblioteca Comunitária/Ponto de Leitura” na tabela 03 acima;

c. **Substantivos ou frases substantivas:** termos normalmente são substantivos + adjetivos ou frases substantivadas. Deve-se evitar verbos e adjetivos sozinhos. Sugere-se aqui uma revisão detalhada dos termos candidatos que forem efetivamente implementados para a ontologia de maneira a evitar termos desses gêneros;
d. **Nomes próprios:** entidades com nomes próprios devem ficar separados numa taxonomia individual de forma a facilitar o controle de identidades;
e. **Qualificadores por parênteses:** uma das formas de lidar com múltiplos termos com significados diferentes em uma taxonomia é utilizando parênteses para especificar significado de cada termo. Exemplo: Francês (língua) e Francês (povo). Recomenda-se escolher um tipo de qualificador único para utilizar entre parênteses, tal como o domínio dos termos ou um tipo de categoria e aplicar por todos os termos;
f. **Acrônimo:** recomenda-se usar acrônimos caso somente eles forem mais conhecidos que as expressões literais. Há vários acrônimos ao longo dos termos descritos, tais como SNIIC, SALIC, SAV, entre outros. Deve-se fazer a avaliação do grau de conhecimento desses acrônimos e deliberar explicitamente pelo seu uso ou expressão literal.

Segundo a norma técnica utilizada, essas são as recomendações e problemas técnicos identificados nos termos candidatos.

6. Considerações

Há 3 aspectos que gostaríamos de recuperar e conectar nas considerações finais deste relatório.

O primeiro deles é a respeito da utilização das redes semânticas como forma de visualização das fontes de informação e identificação dos termos candidatos mais importantes em relação a estrutura da rede. A rede nos permite identificar as interconexões entre as fontes de informação, os termos sobrepostos, os termos conectores e as posições estruturais ocupadas por eles, além de podermos visualizar o efeito na rede quando termos entram e saem. Entendemos que esse recurso deve ser utilizado continuamente no processo de construção da ontologia, bem como para futuras análises de novas fontes de informação que forem identificadas pelo MinC.

O segundo aspecto diz respeito às normas técnicas. Entendemos que elas fornecem indicações que ajudam a normalizar os termos e facilitar sua compatibilização com outras iniciativas que adotam essas normas como parâmetros de construção de termos. Sugerimos que todos os termos sejam analisados segundo essas normas e possam ser ajustados conforme suas indicações.

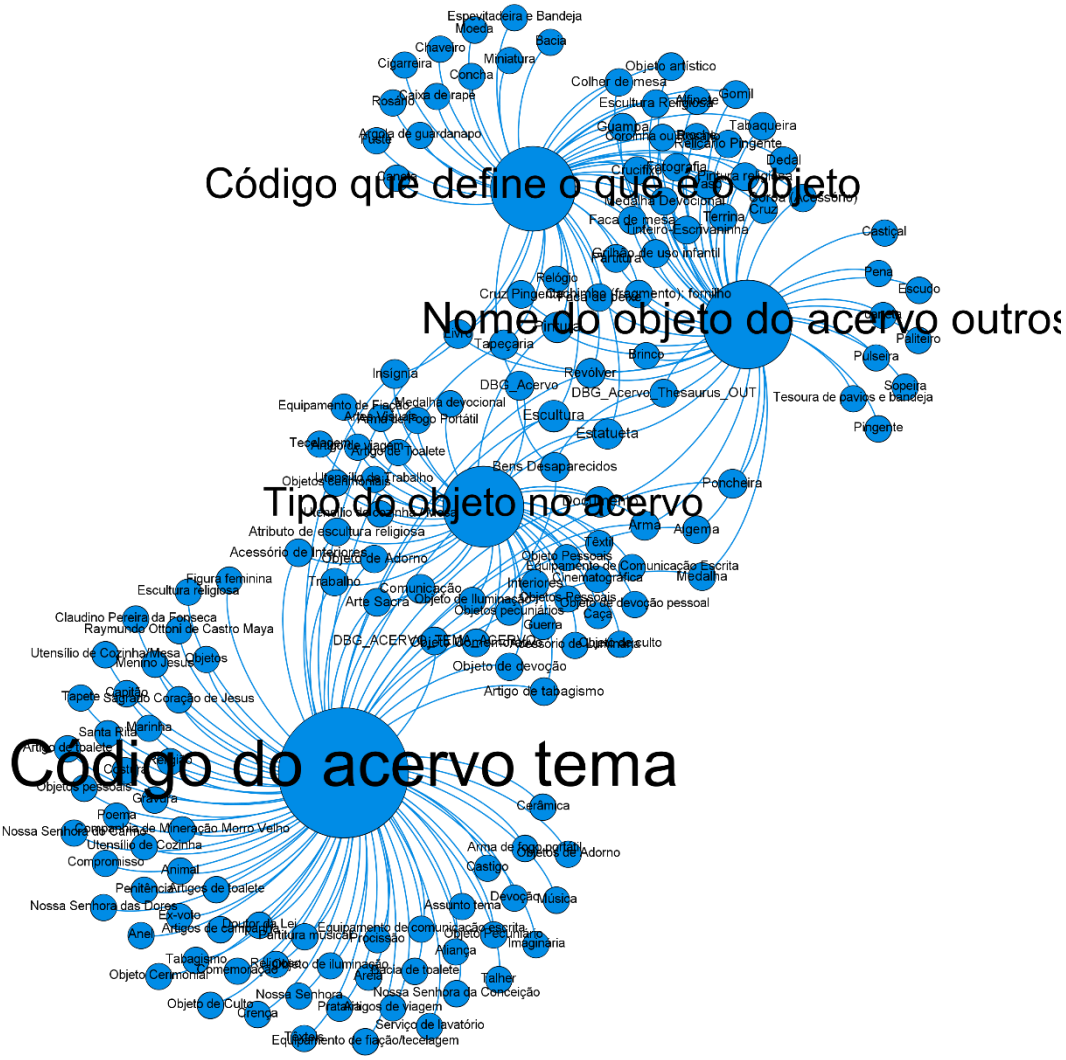
O terceiro diz respeito a como esse estudo aponta possibilidades de análise de termos que sugerimos serem incorporadas no processo de construção da ontologia.

Priorizar os termos de maior centralidade, bem como as fontes relacionadas ao “Mapas Culturais”, “Equipamentos culturais” e “SALIC” como foco prioritário para a ontologia. Vale reforçar que a centralidade desses termos e dessas fontes ocupa uma posição estratégica para o MinC como meio de facilitar o diálogo com outras áreas e favorecer sua interconectividade.

Apêndice 01
Redes Semânticas por fonte de informação

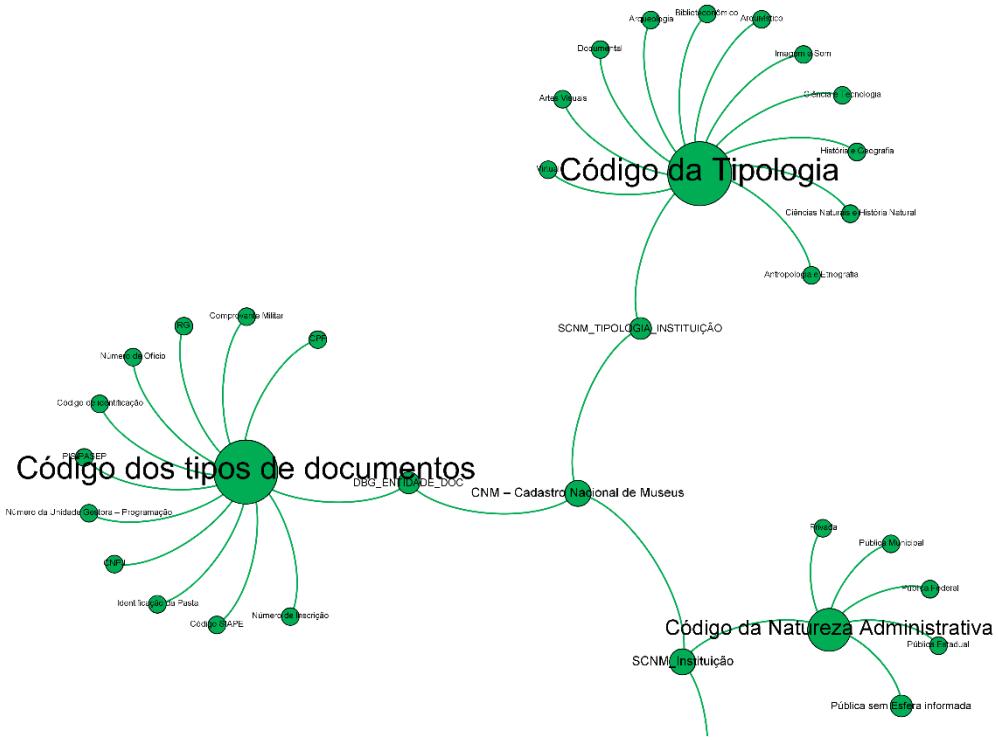
IBRAM

Figura 04 – Bens desaparecidos



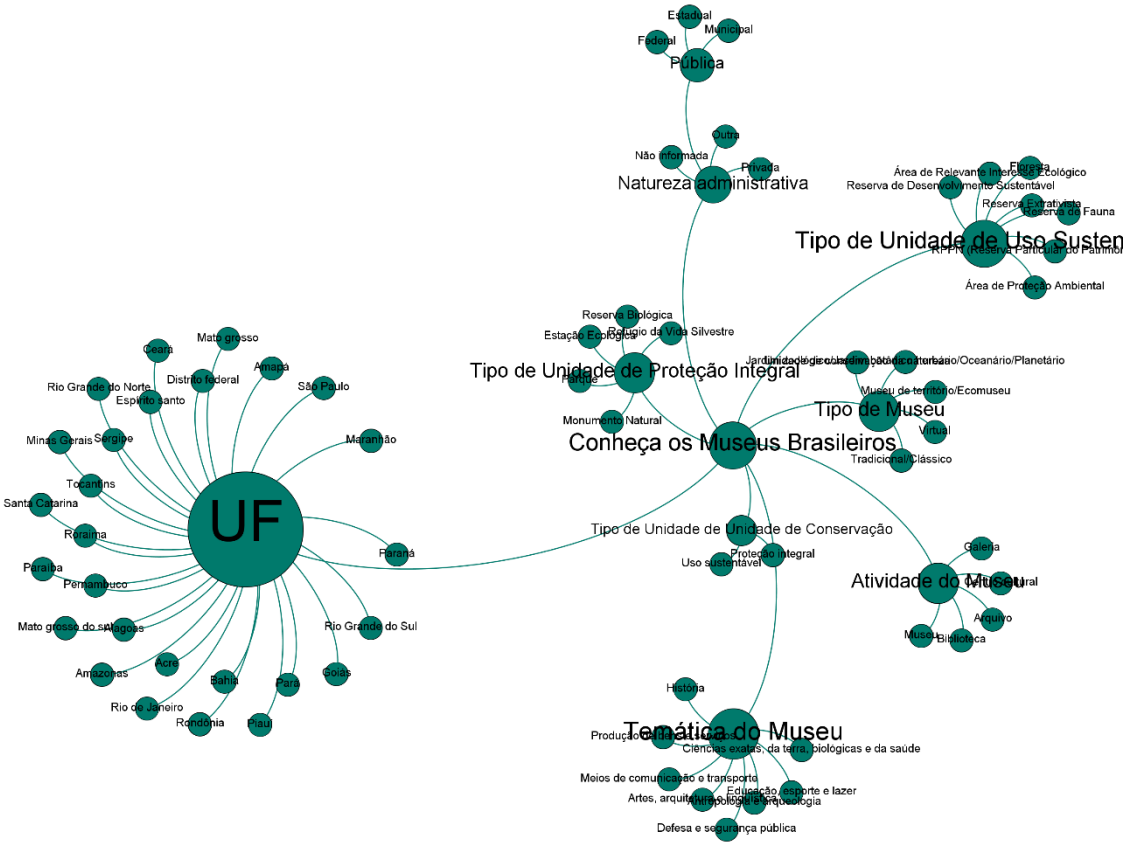
Fonte: dados da pesquisa.

Figura 05 – Cadastro nacional de museus



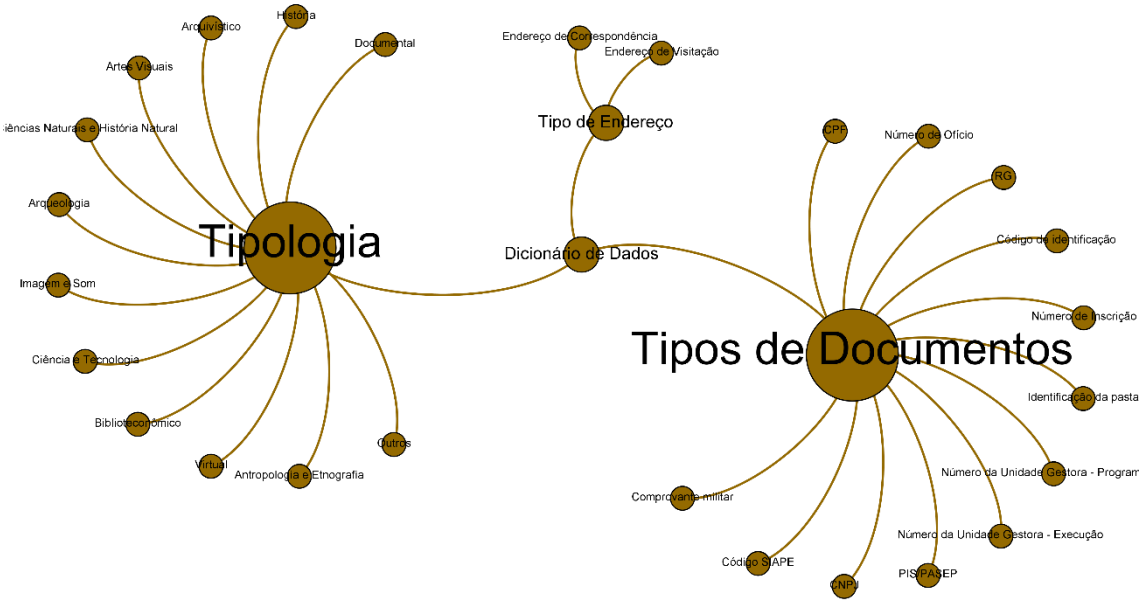
Fonte: dados da pesquisa.

Figura 06 – Conheça os museus brasileiros



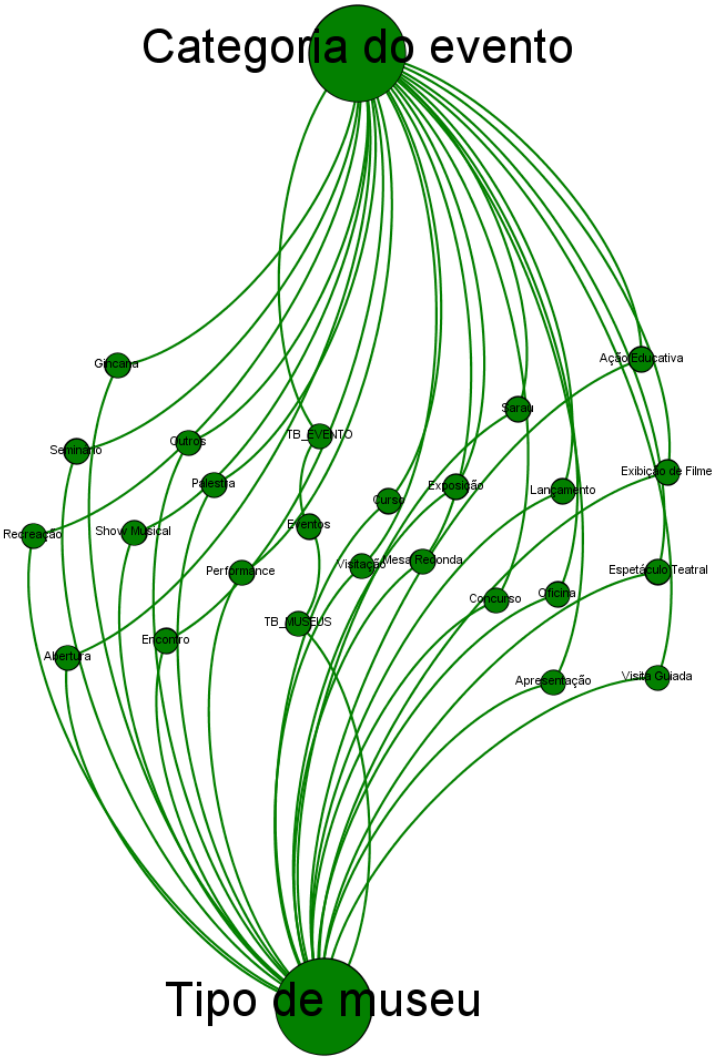
Fonte: dados da pesquisa.

Figura 07 – Dicionário de dados



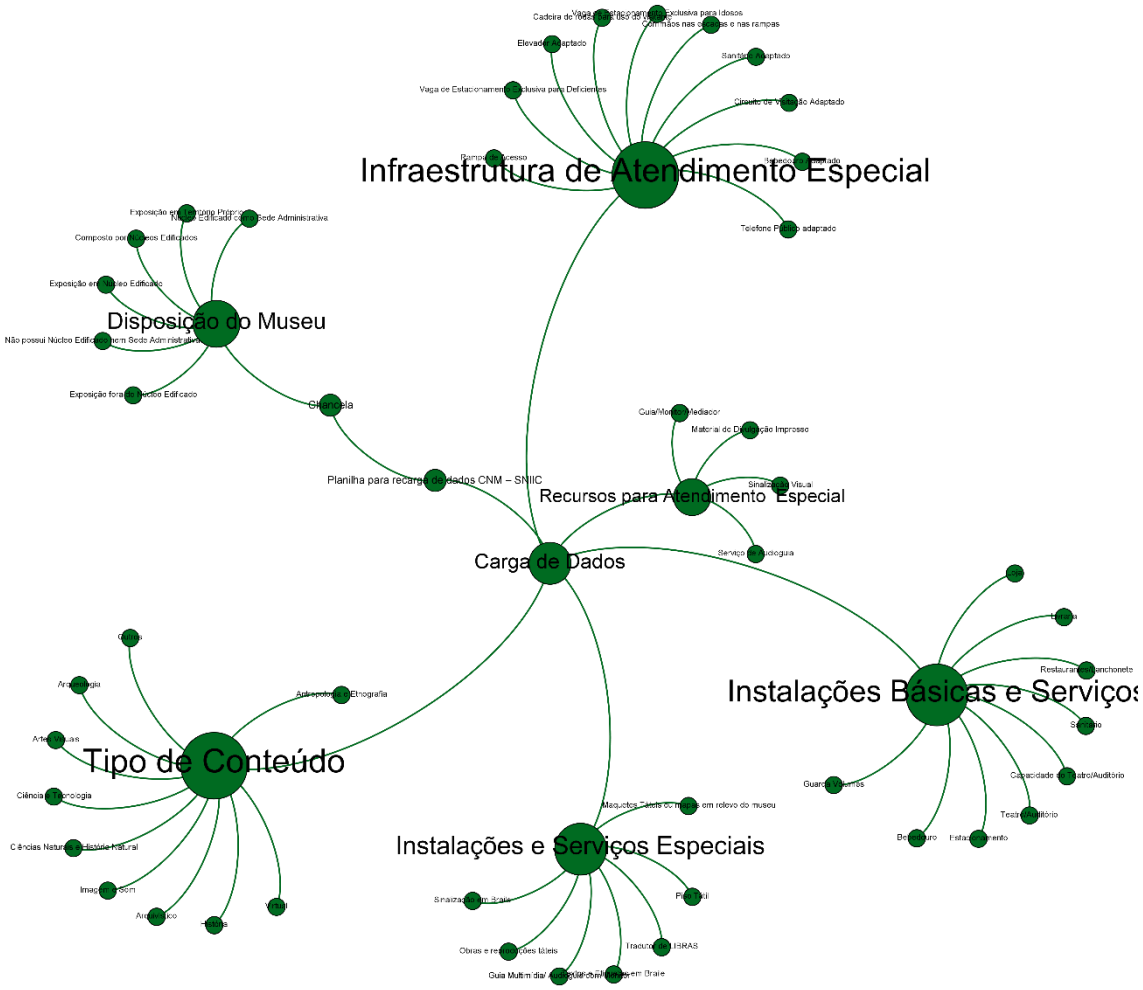
Fonte: dados da pesquisa.

Figura 08 - Eventos



Fonte: dados da pesquisa.

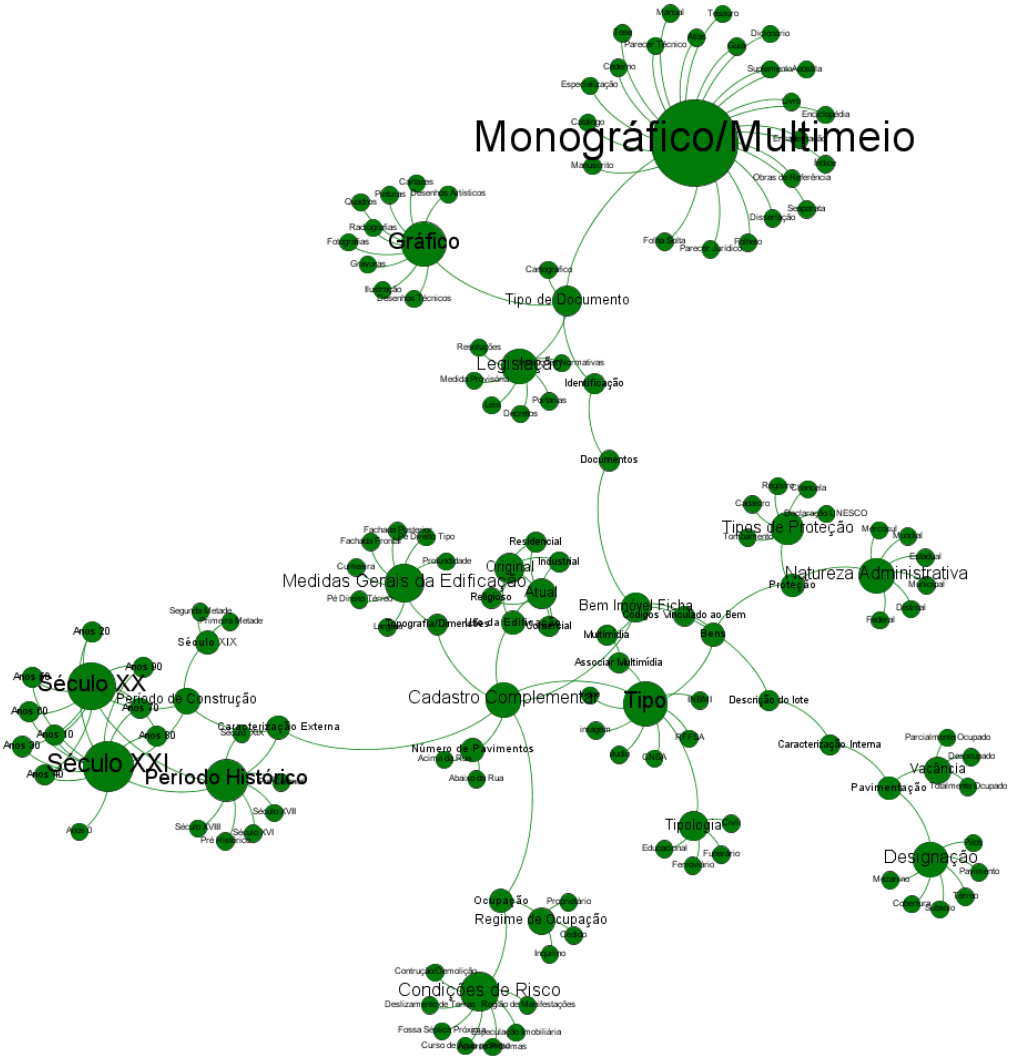
Figura 09 – Planilha de recarga SNM - SNIIC



Fonte: dados da pesquisa.

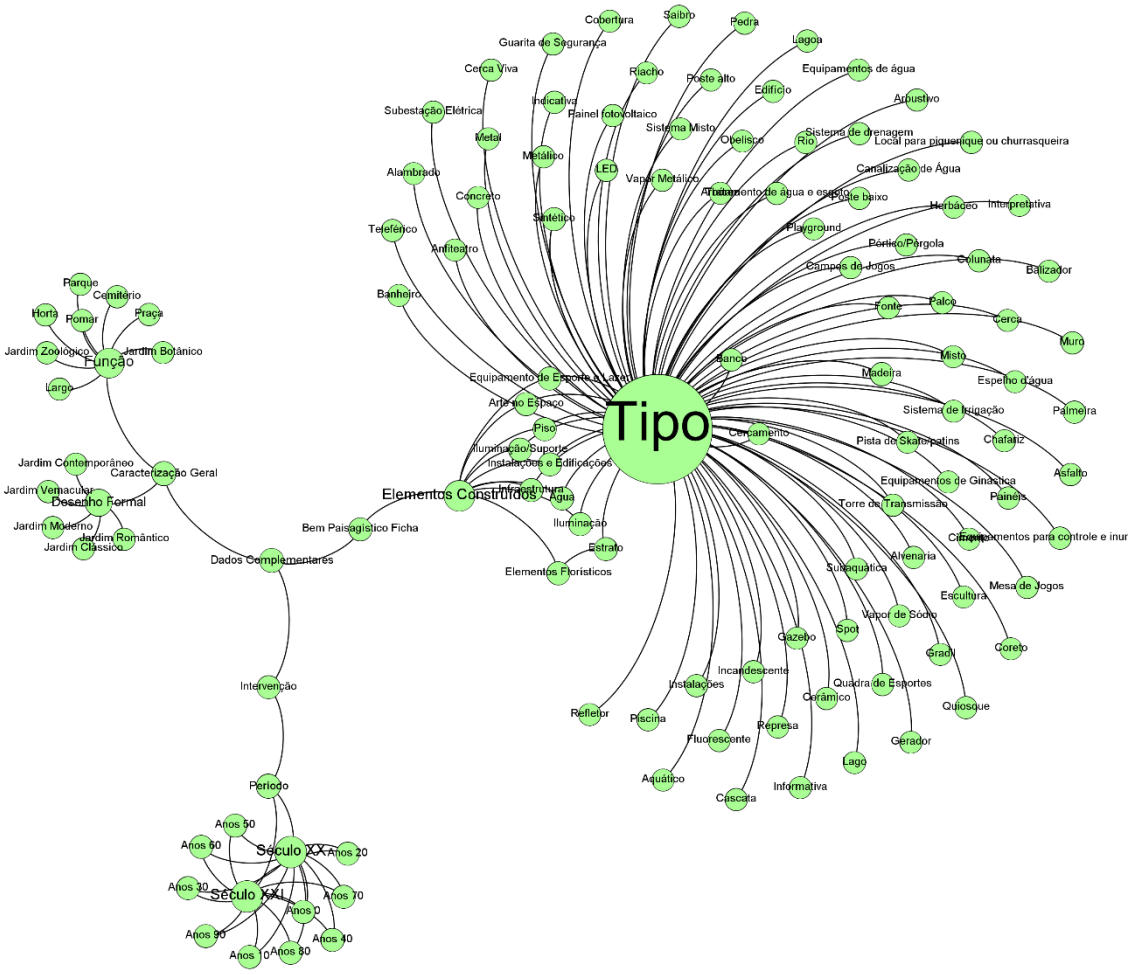
IPHAN

Figura 10 – Ficha Bem imóvel



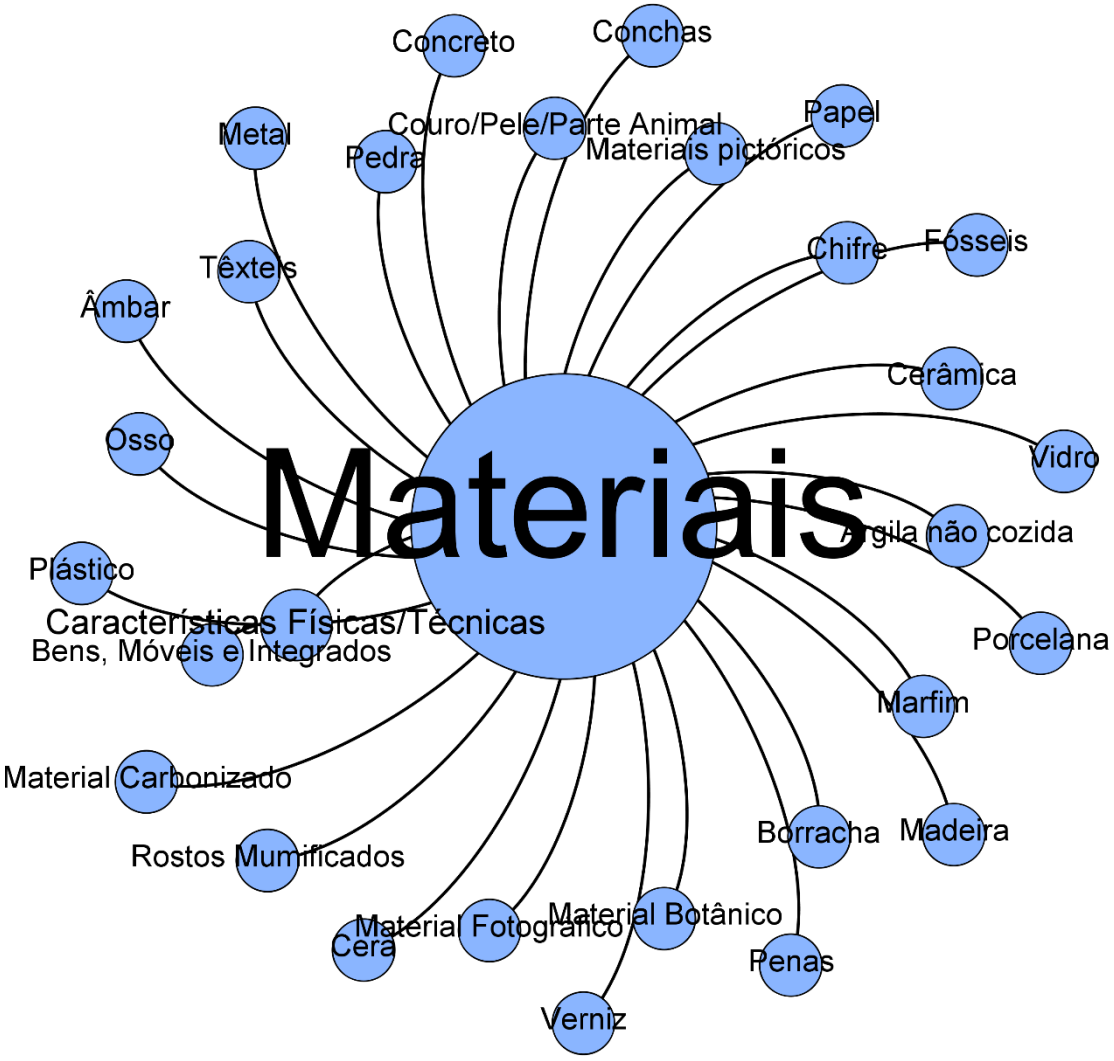
Fonte: dados da pesquisa.

Figura 11 – Ficha Bem paisagístico



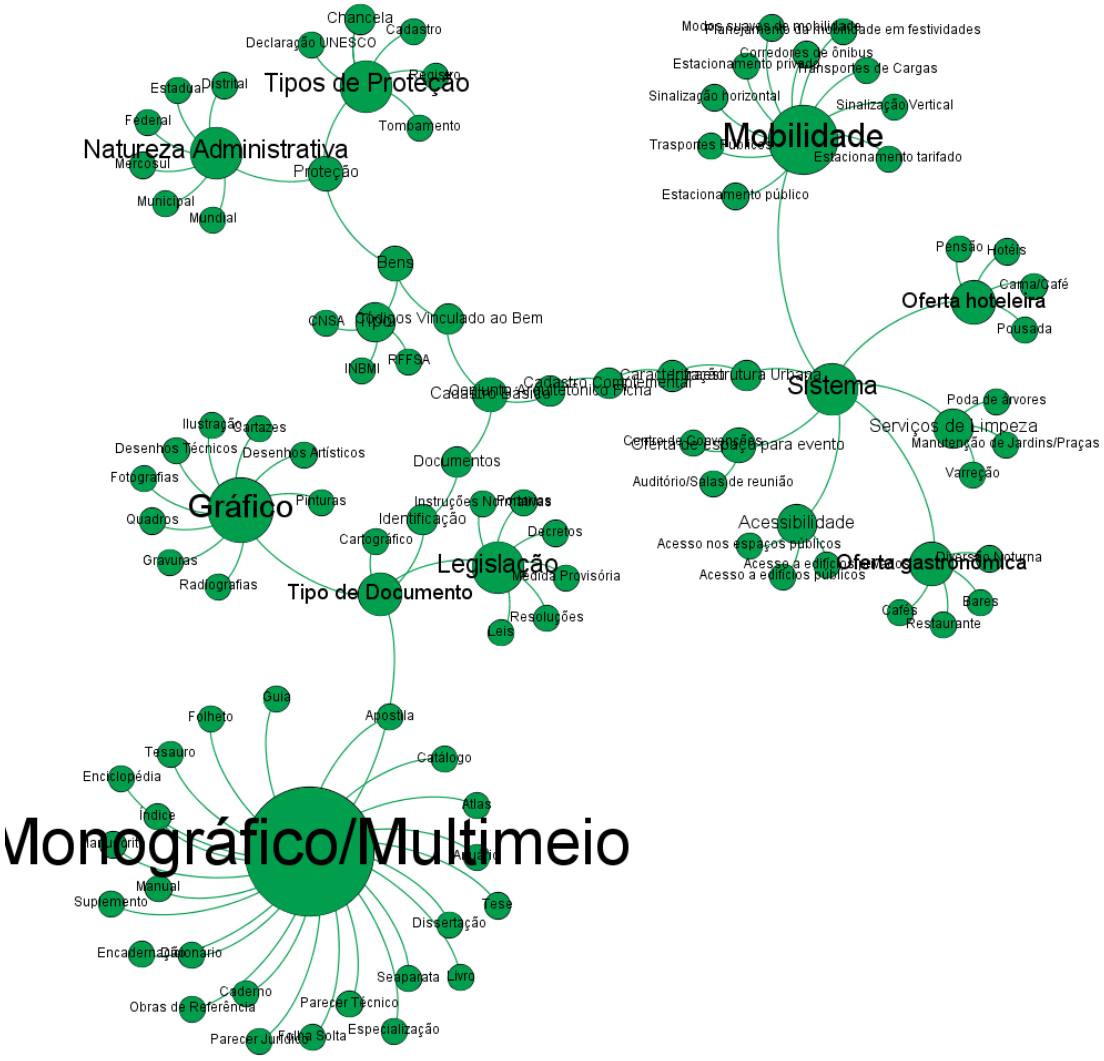
Fonte: dados da pesquisa.

Figura 12 – Bens, móveis e integrados



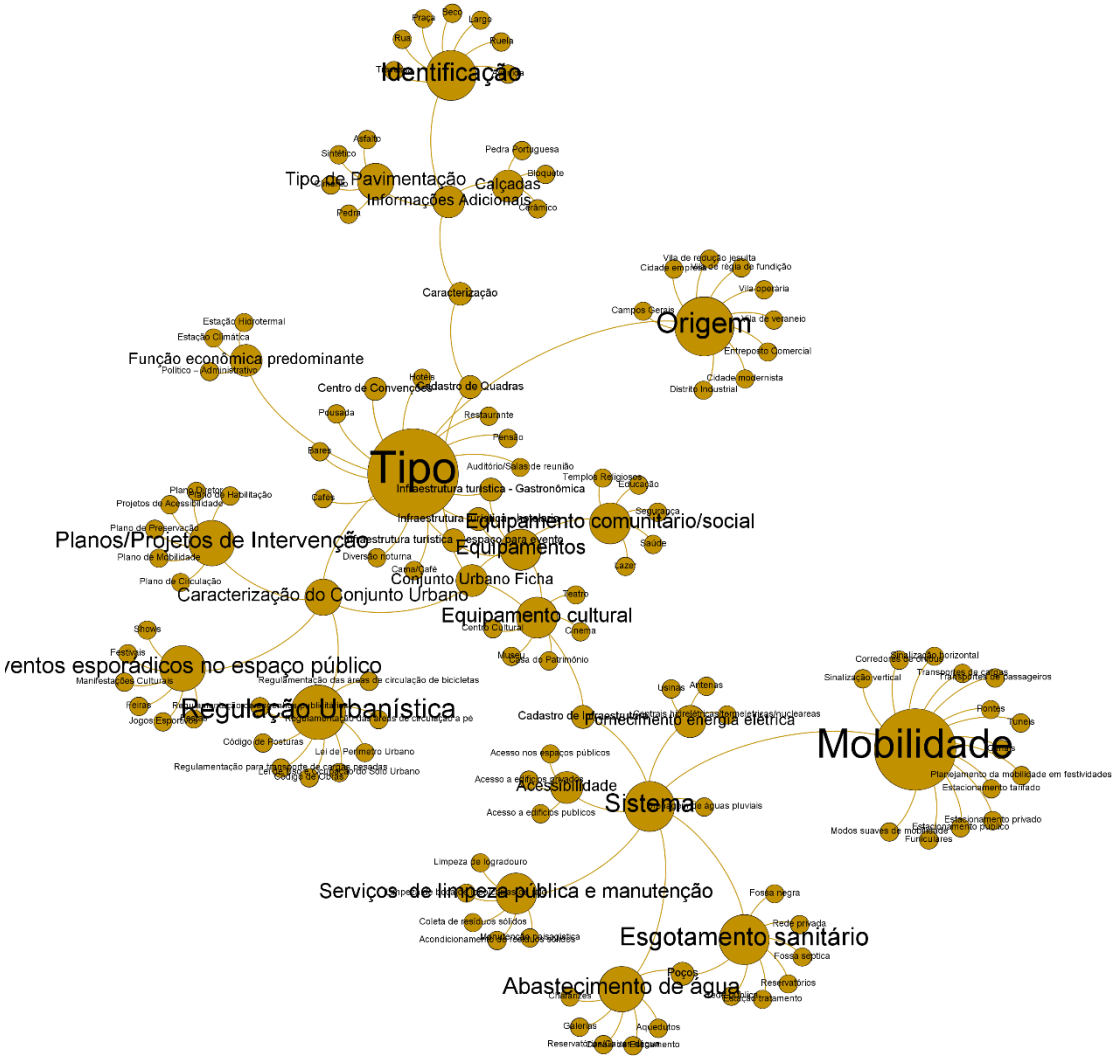
Fonte: dados da pesquisa.

Figura 13 – Ficha Conjunto arquitetônico



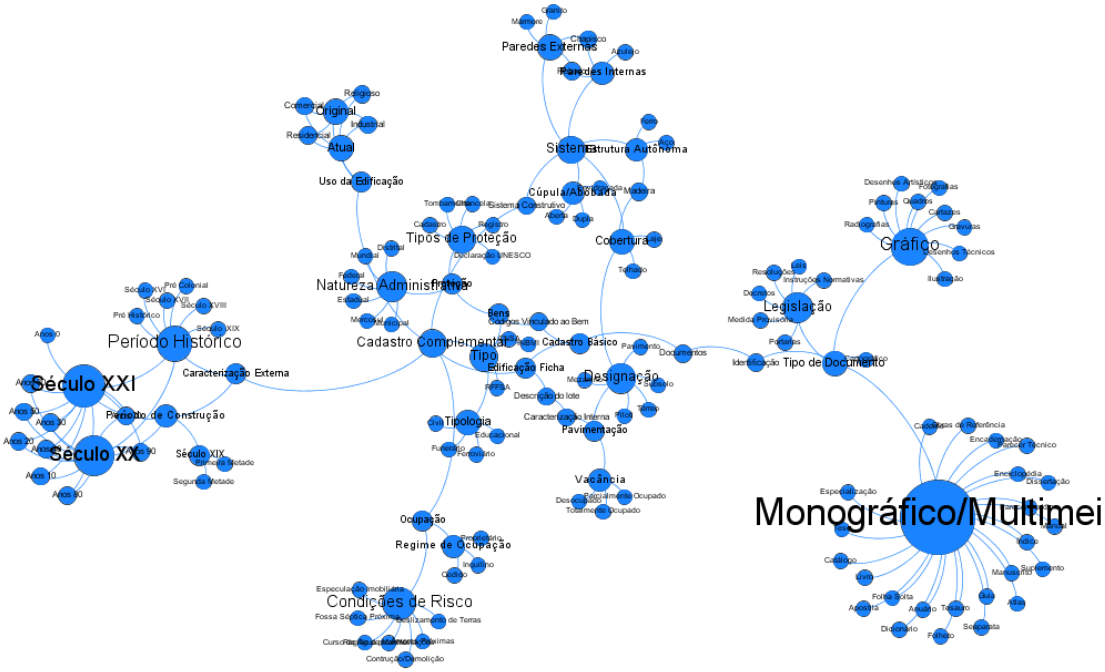
Fonte: dados da pesquisa.

Figura 14 – Ficha Conjunto urbano



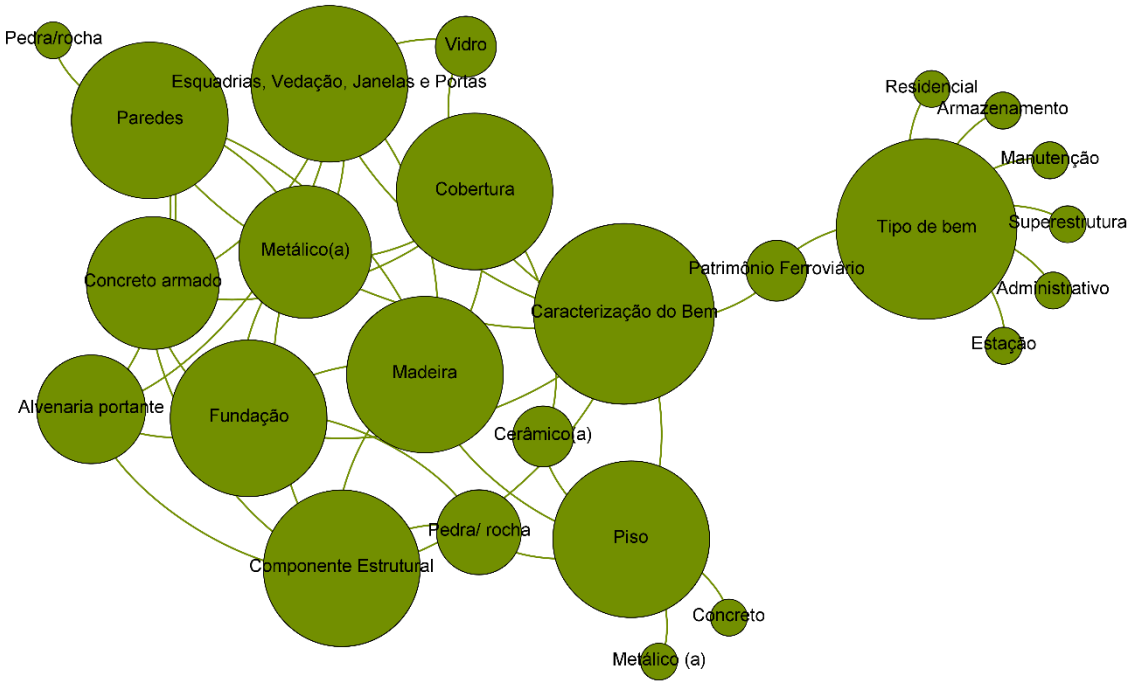
Fonte: dados da pesquisa.

Figura 15 – Ficha Edificação



Fonte: dados da pesquisa.

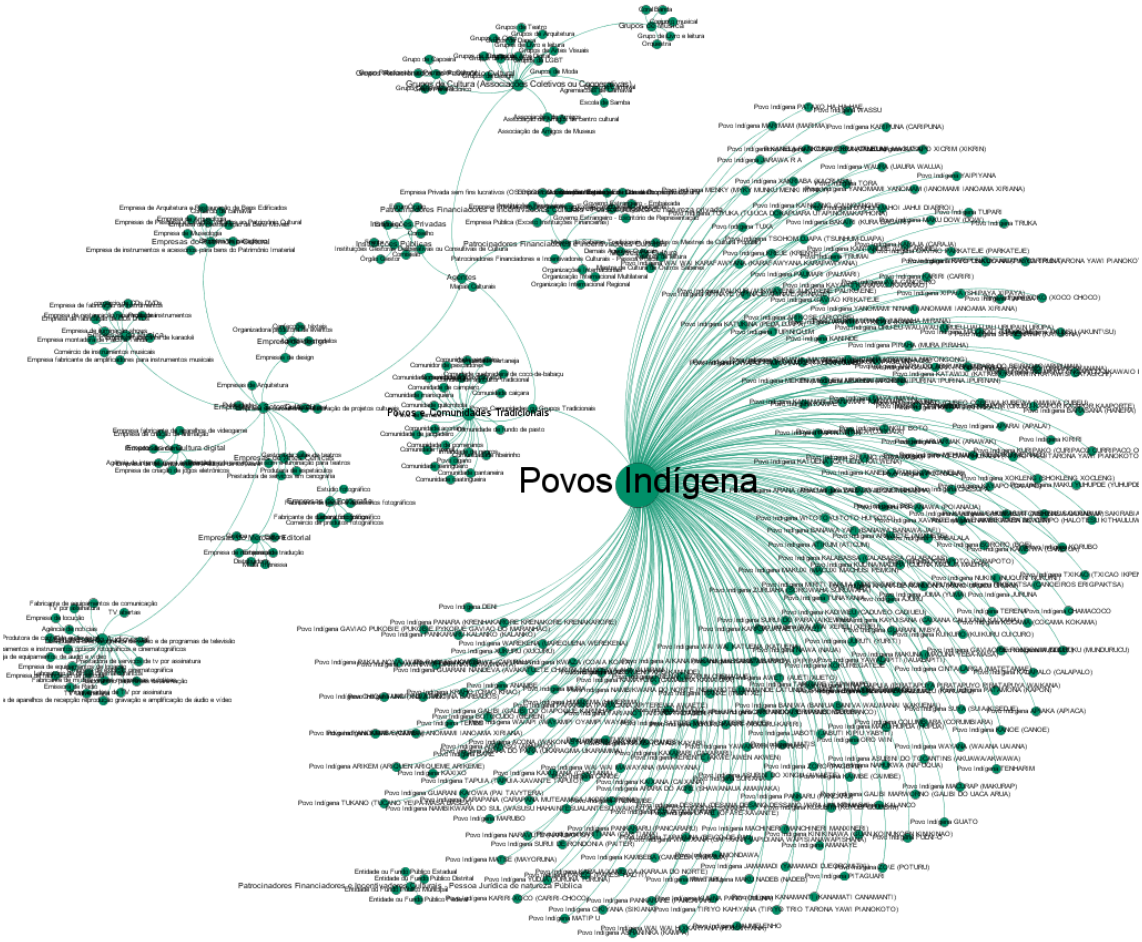
Figura 16 – Patrimônio Ferroviário



Fonte: dados da pesquisa.

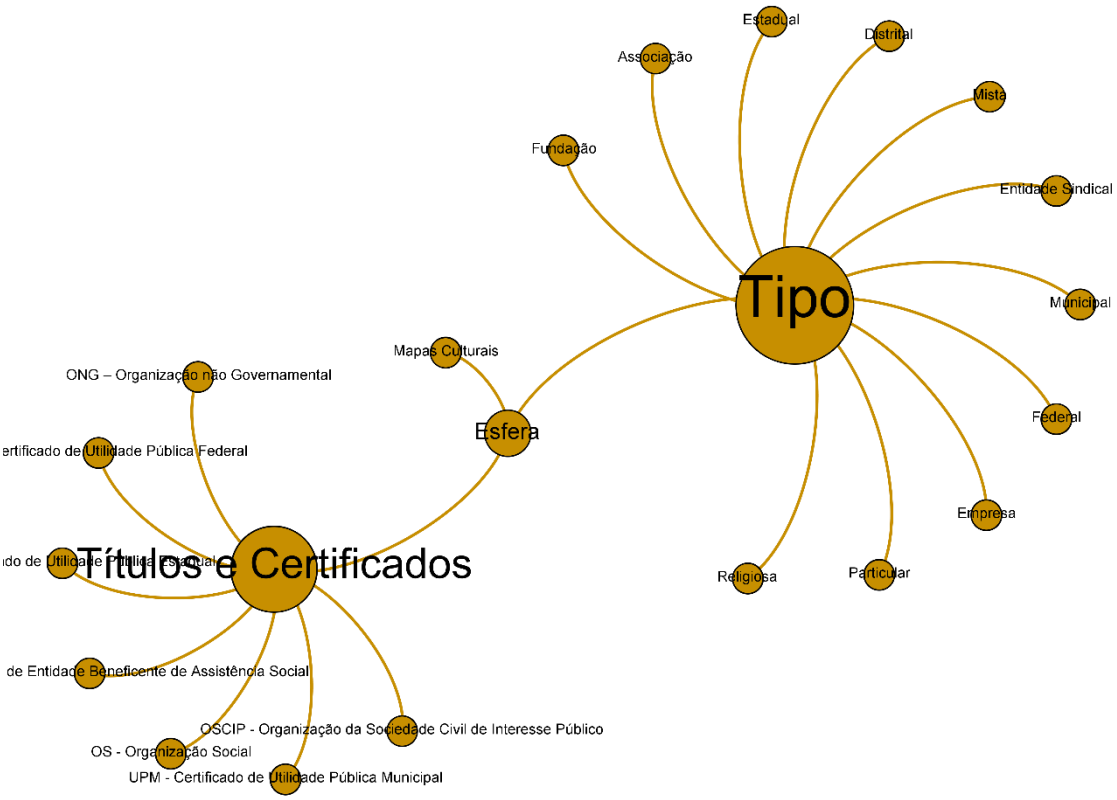
Mapas Culturais

Figura 17 – Agentes



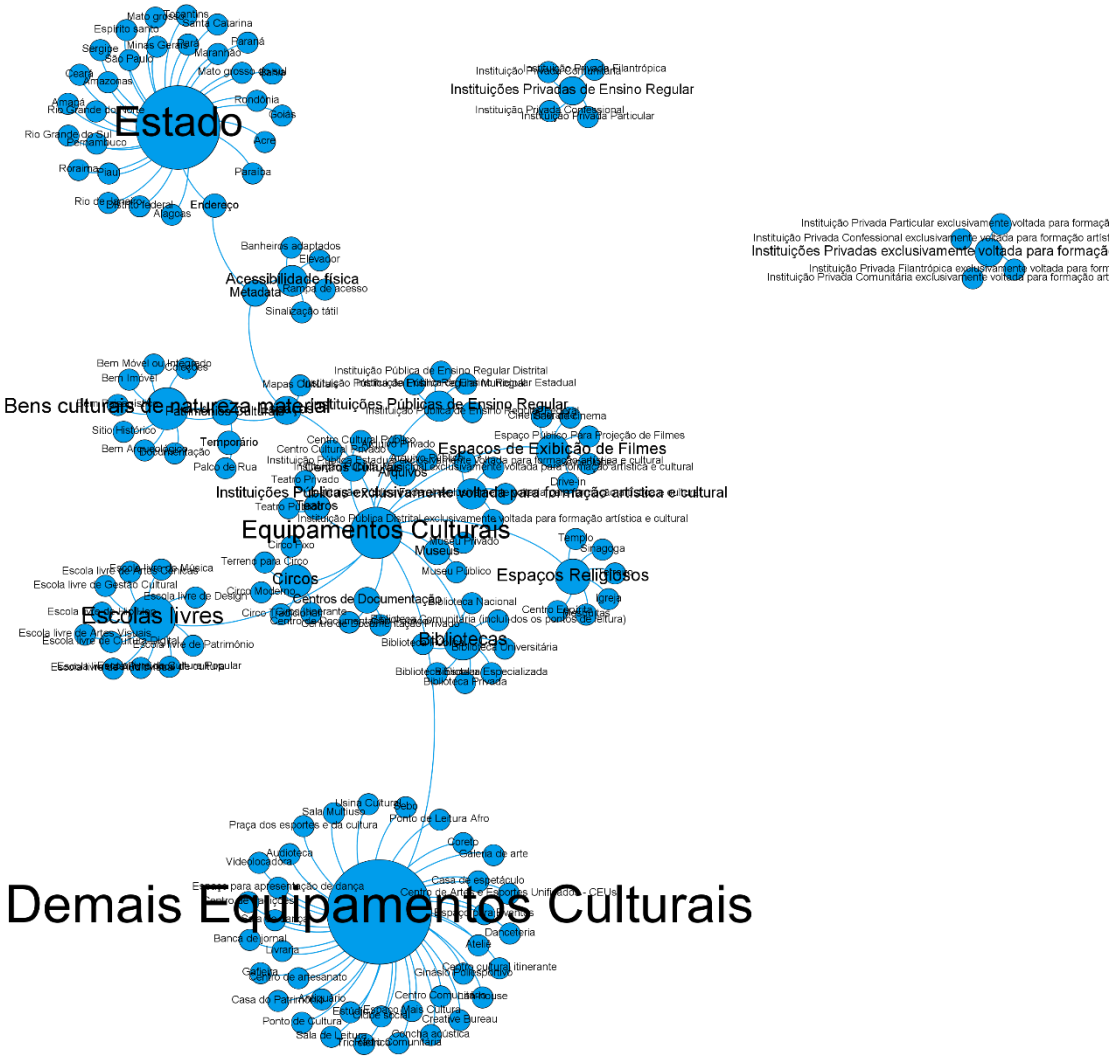
Fonte: dados da pesquisa.

Figura 18 – Esferas



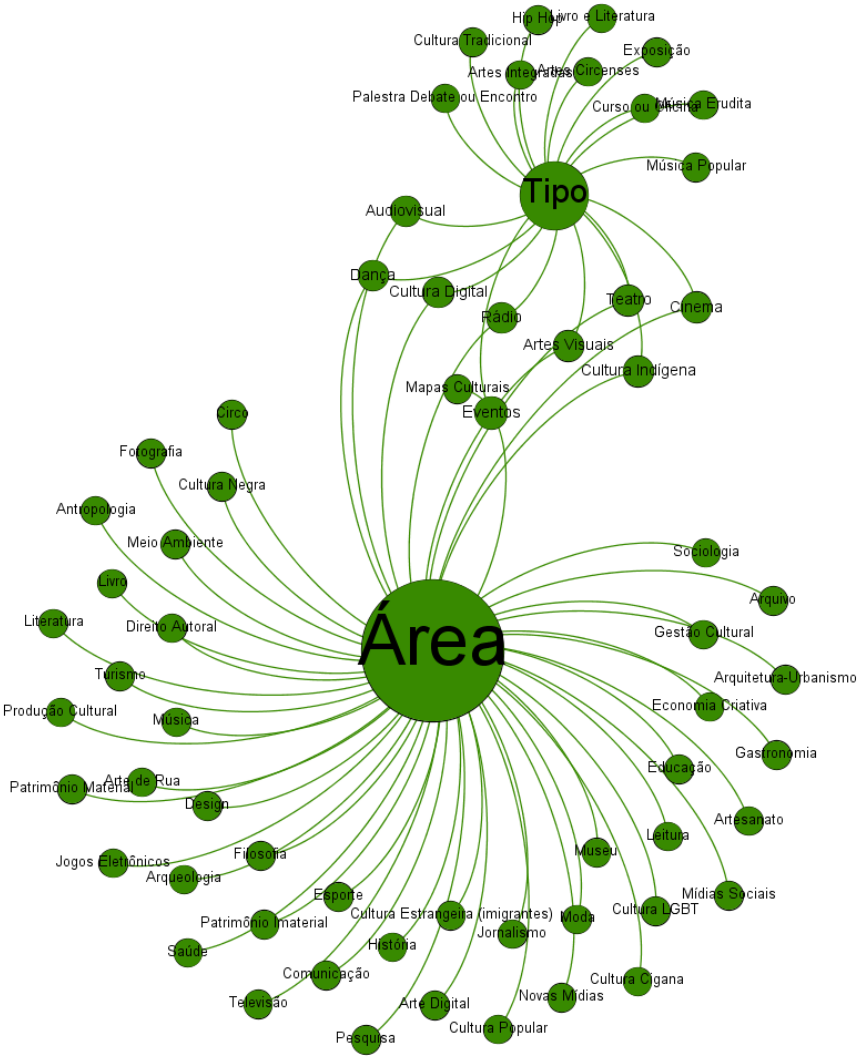
Fonte: dados da pesquisa.

Figura 19 – Espaços



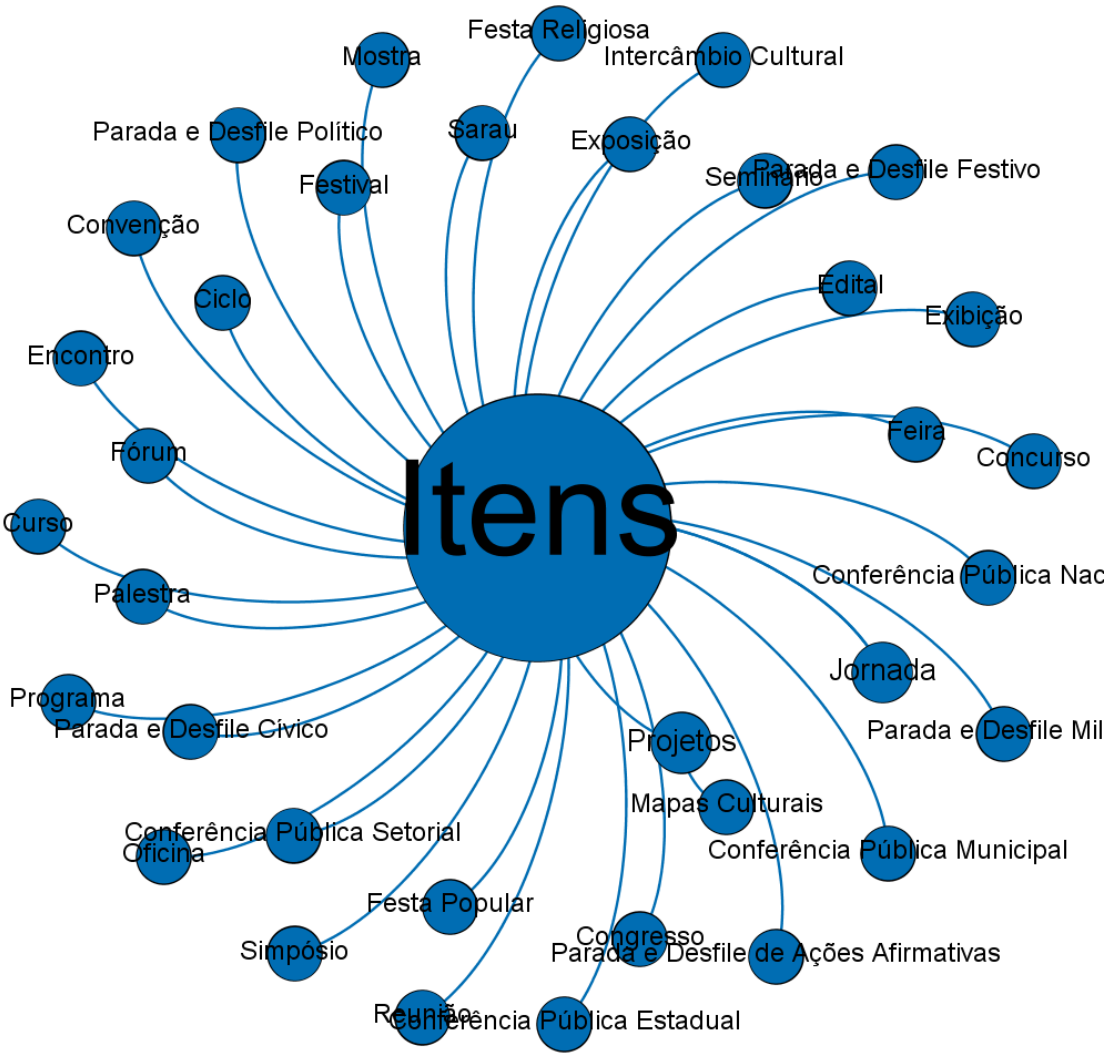
Fonte: dados da pesquisa.

Figura 19 - Eventos



Fonte: dados da pesquisa.

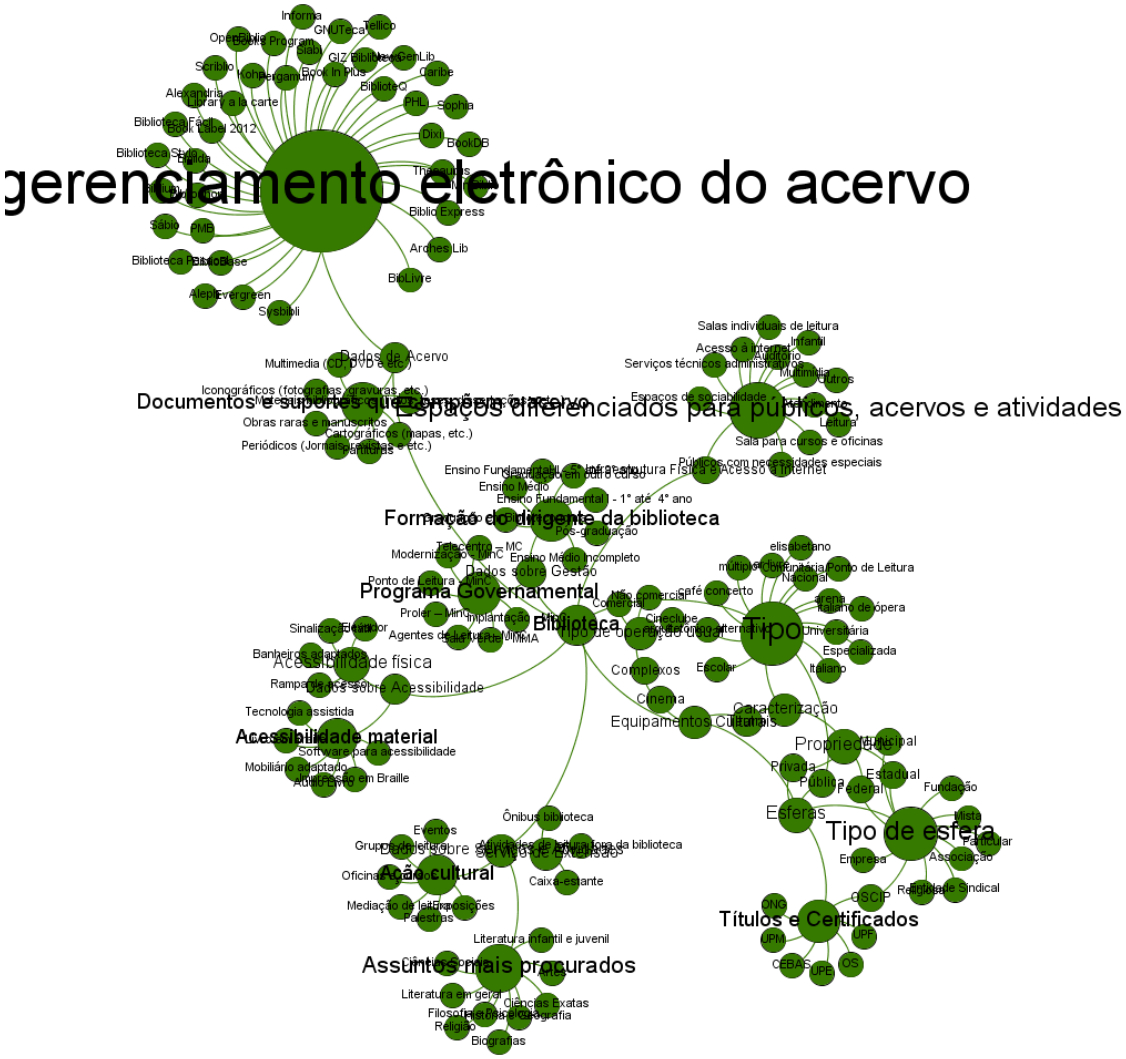
Figura 20 – Projetos



Fonte: dados da pesquisa.

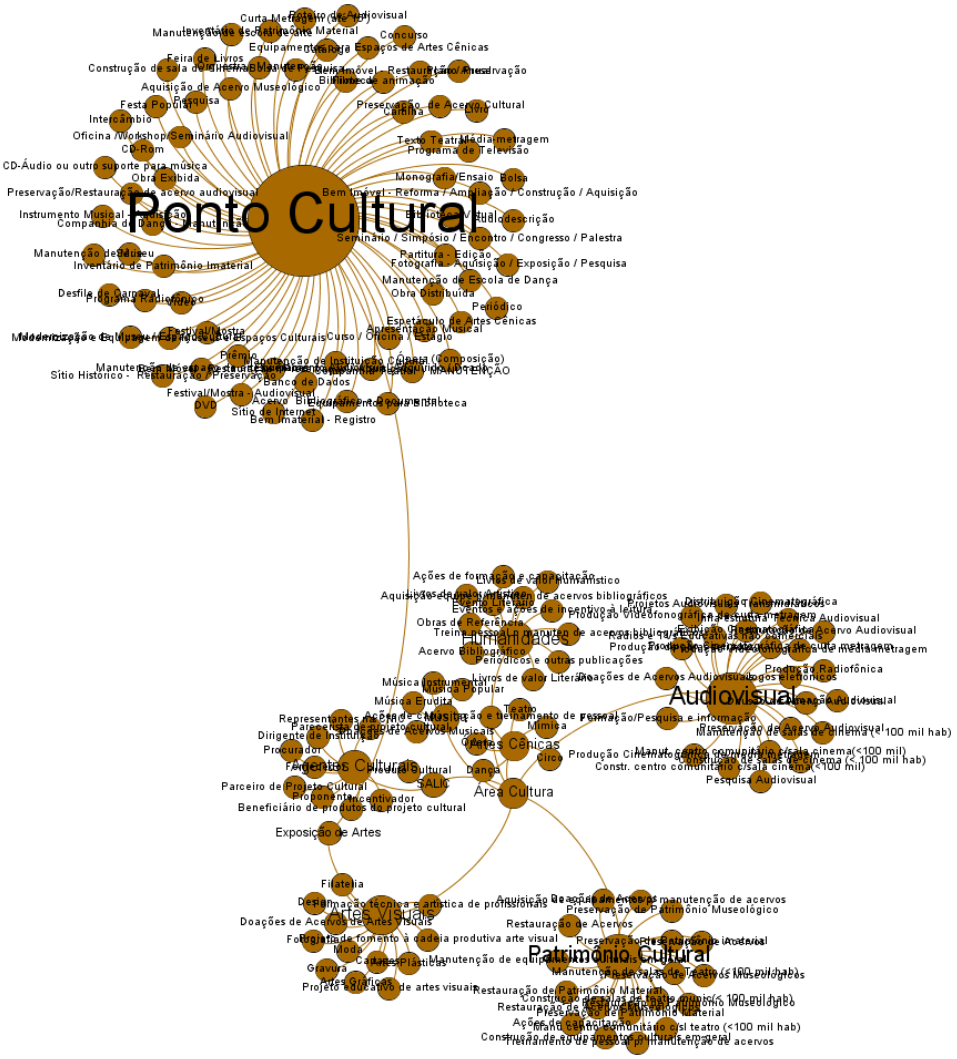
SALIC

Figura 21 – Equipamentos culturais



Fonte: dados da pesquisa.

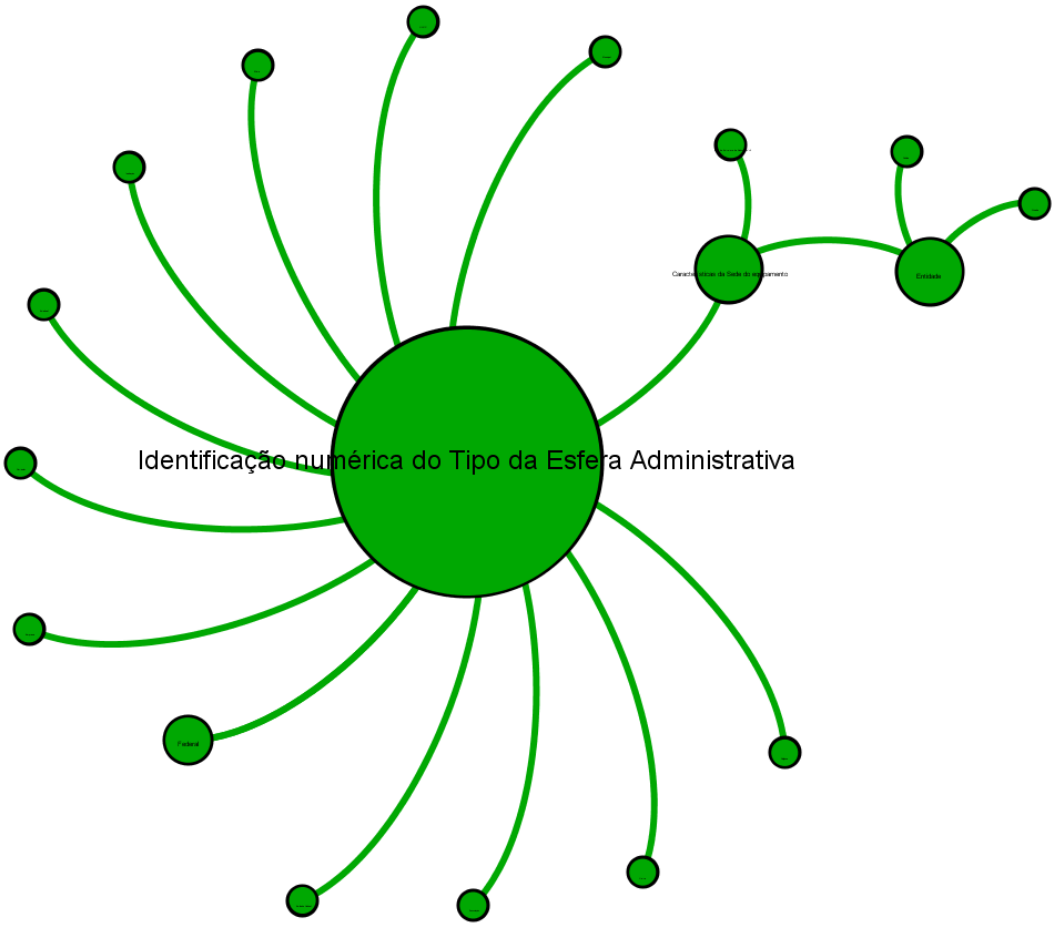
Figura 22 -



Fonte: dados da pesquisa.

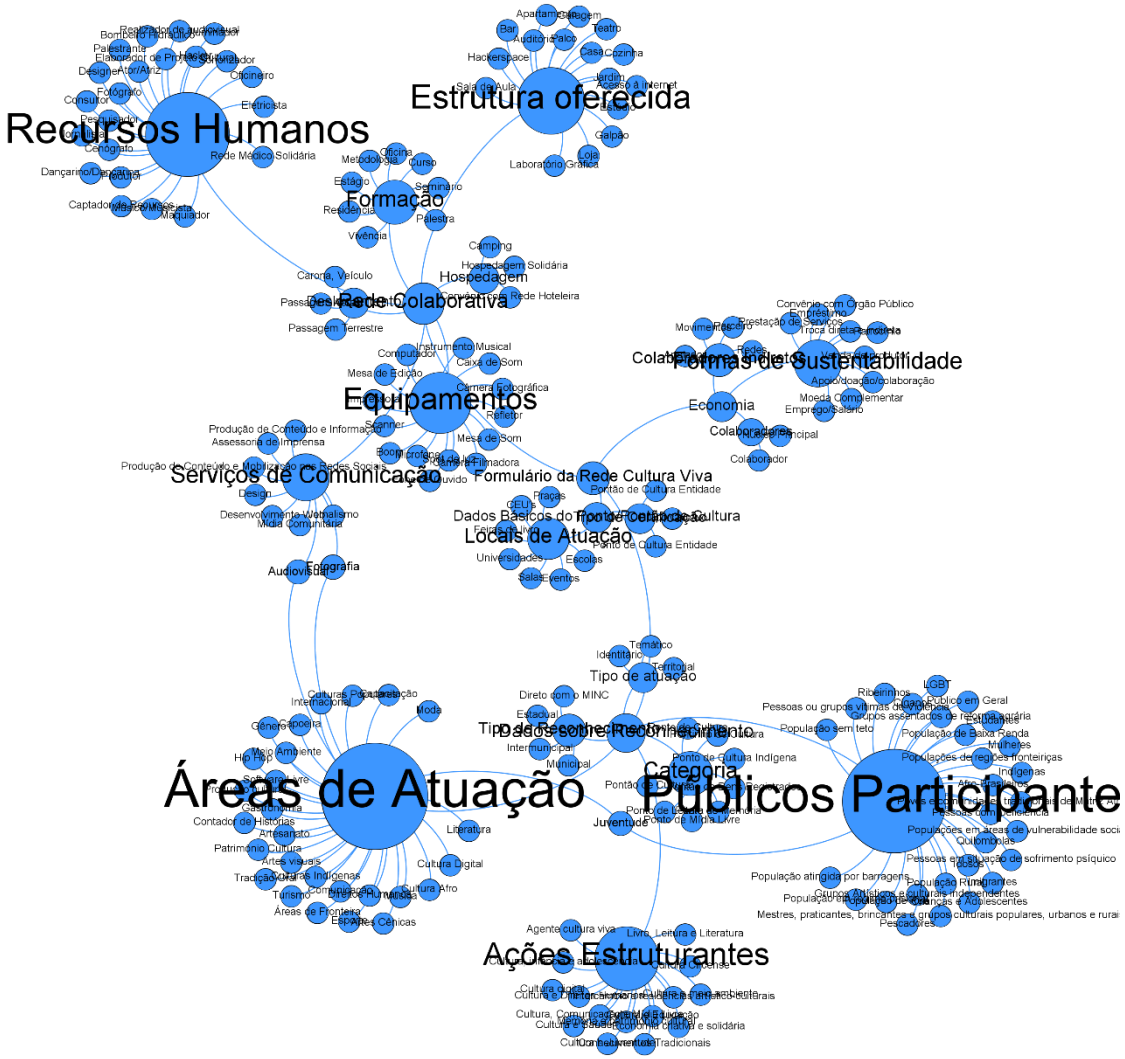
Equipamentos culturais

Figura 23 – Base de estrutura de dados da SAV



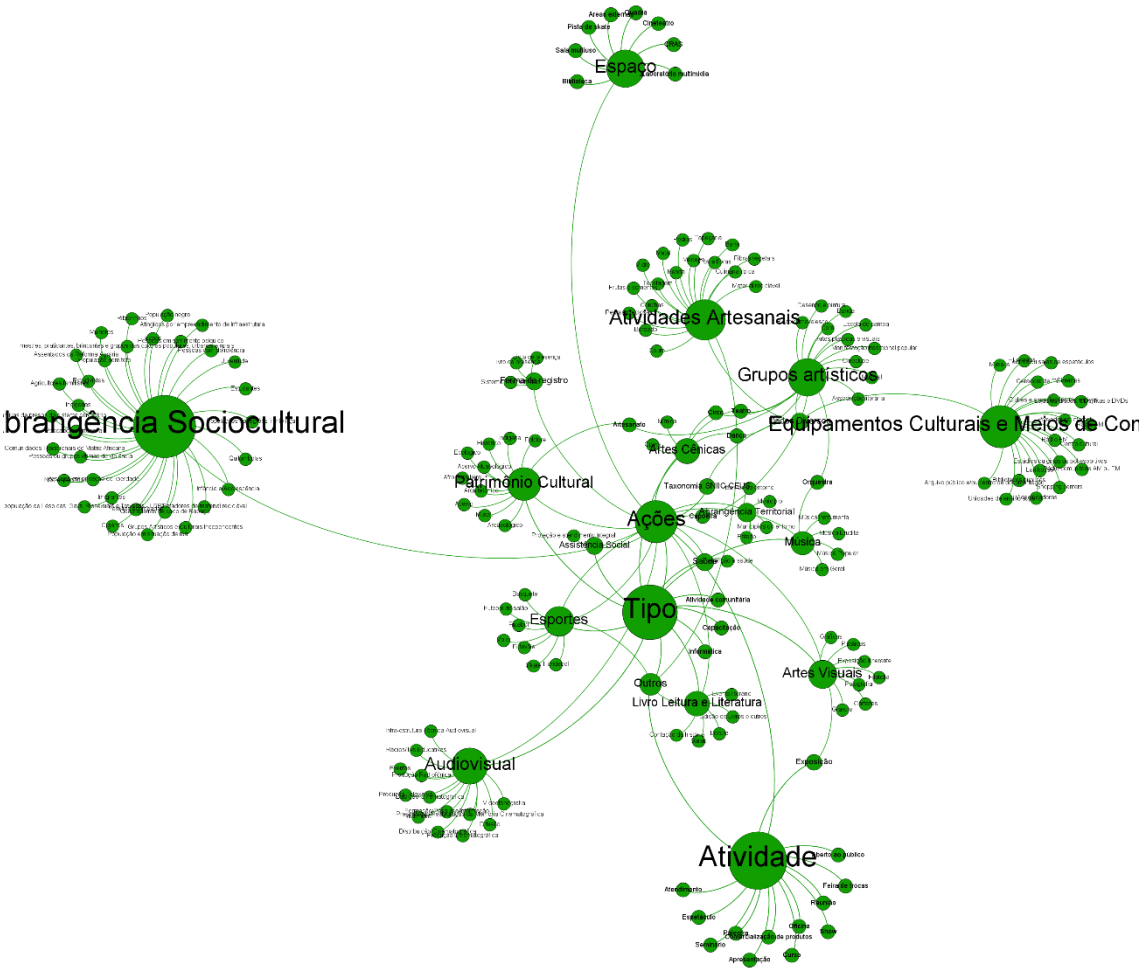
Fonte: dados da pesquisa.

Figura 24 – Formulário Rede Cultura Viva



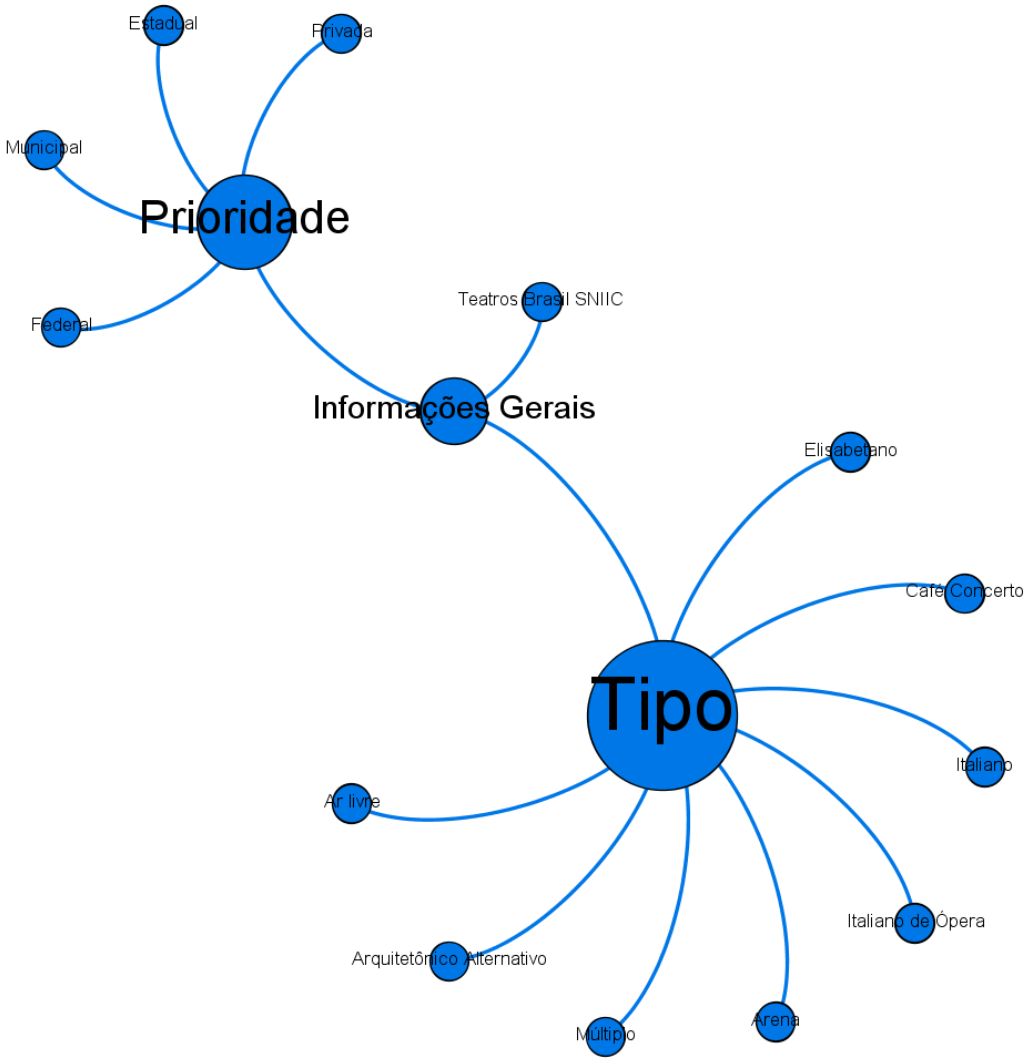
Fonte: dados da pesquisa.

Figura 25 – Taxonomia Sniic - CEUs



Fonte: dados da pesquisa.

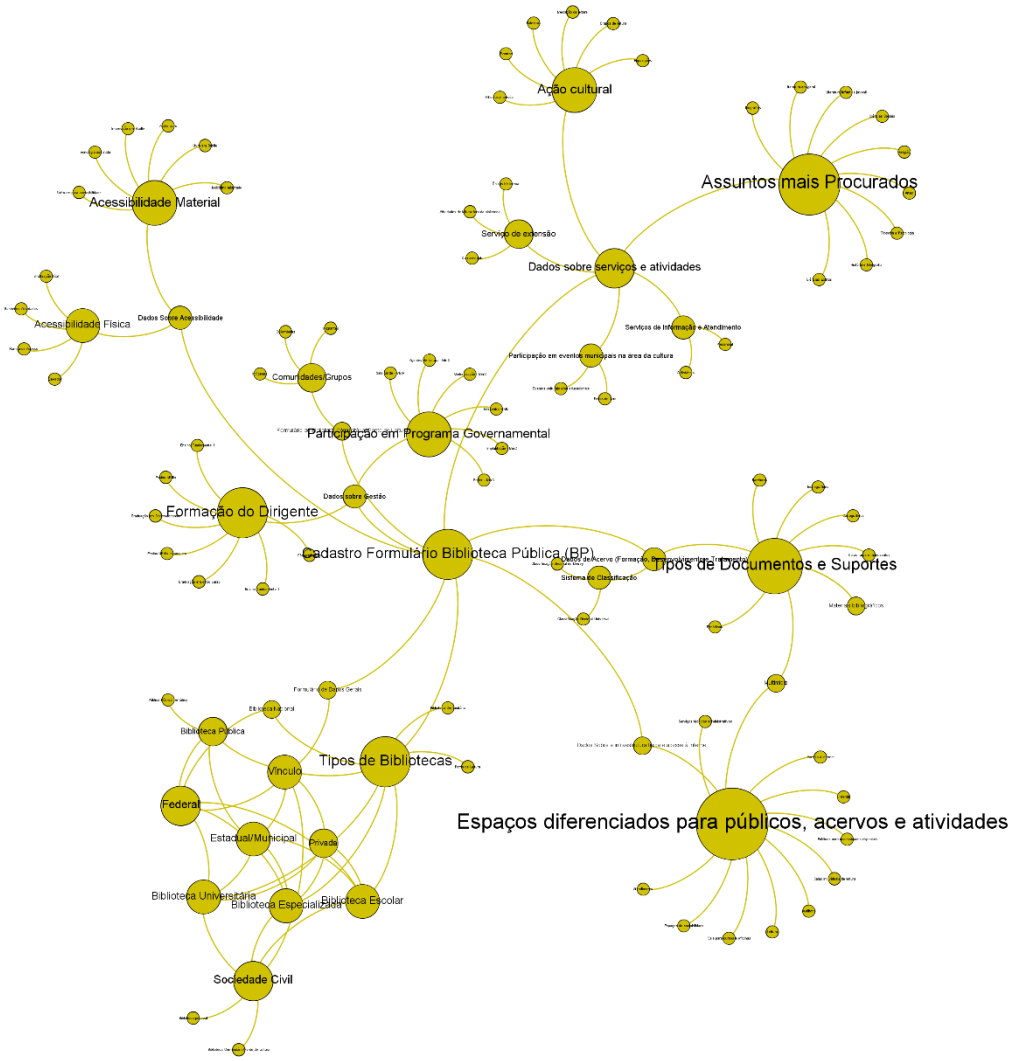
Figura 26 – Teatros Brasil SNIIC



Fonte: dados da pesquisa.

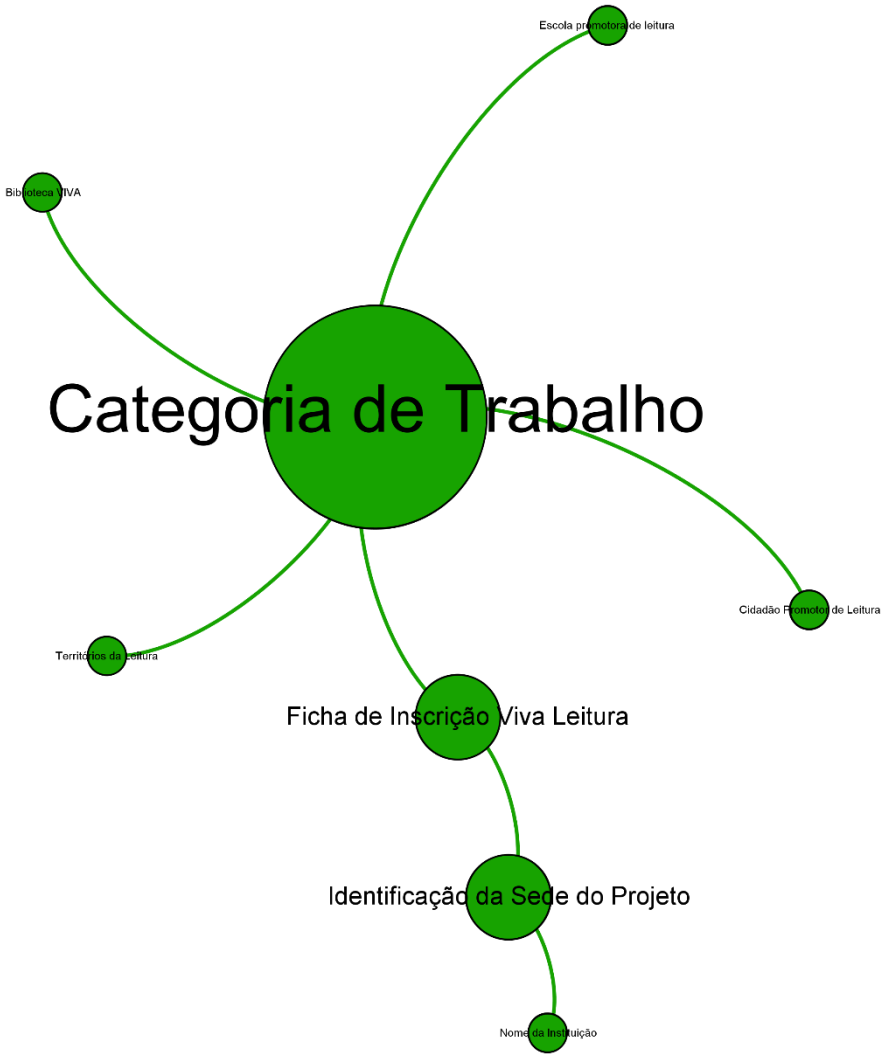
DLLLB

Figura 27 – Cadastro Formulário BP



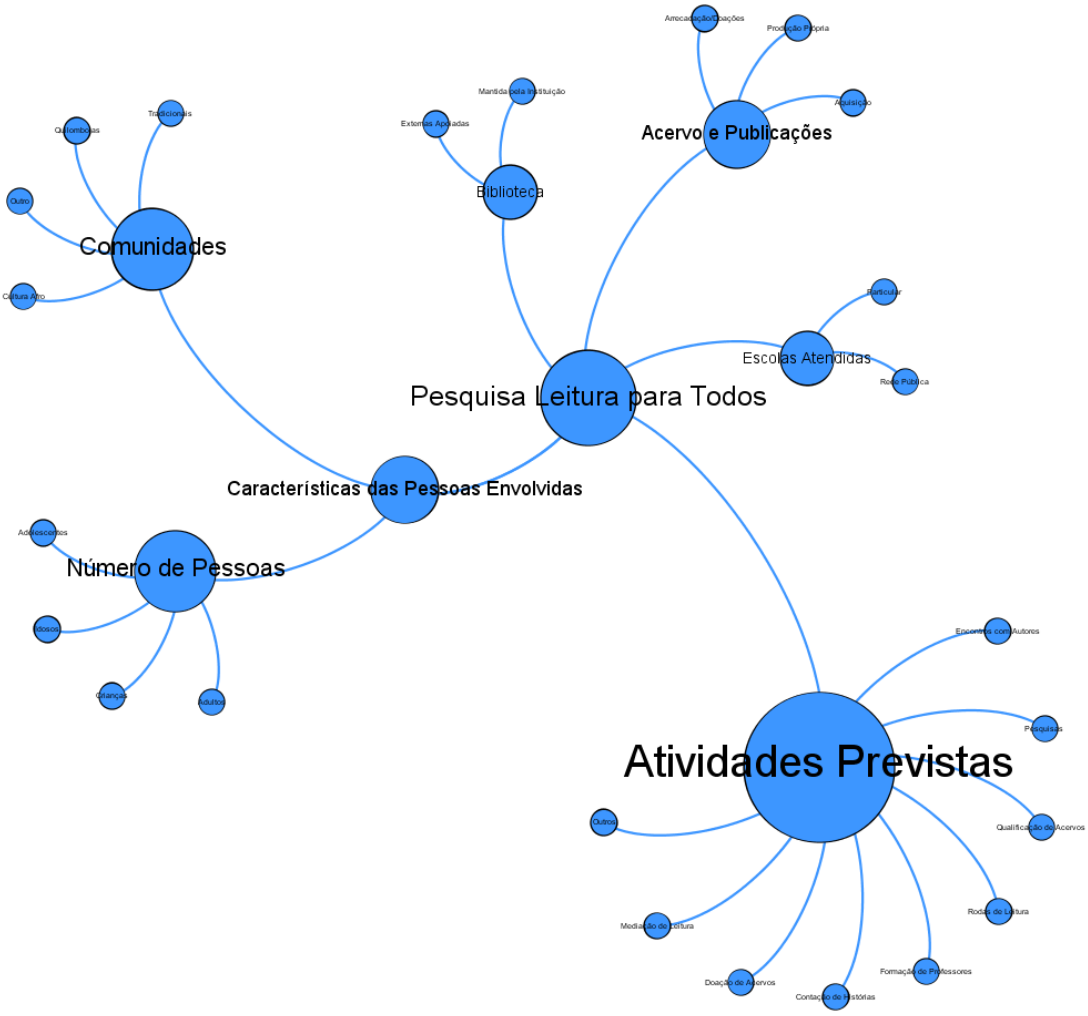
Fonte: dados da pesquisa.

Figura 28 – Ficha de inscrição Leitura Viva



Fonte: dados da pesquisa.

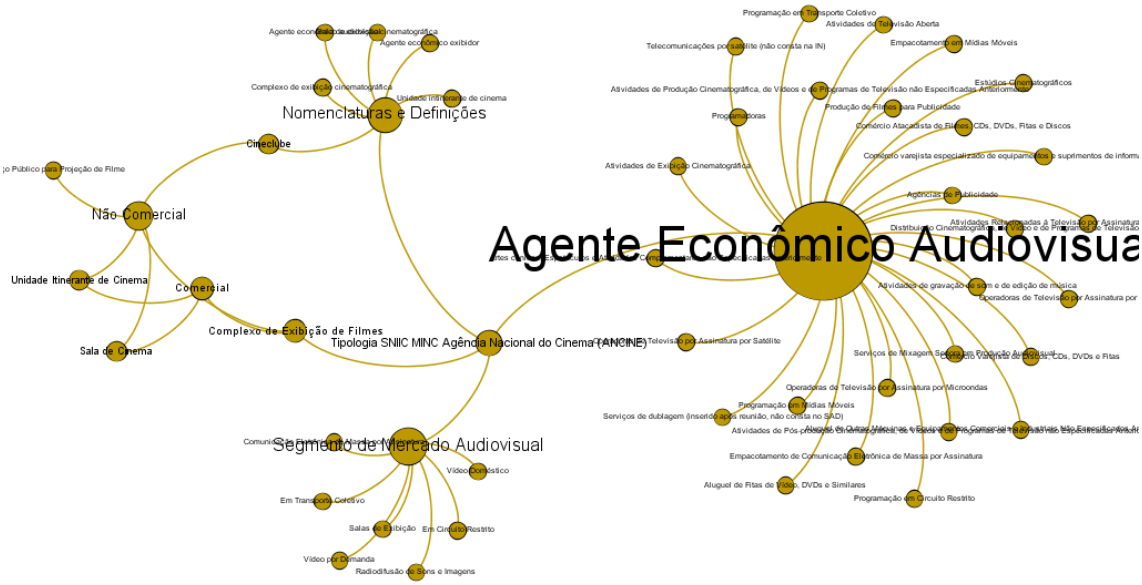
Figura 29 – Pesquisa Leitura para Todos



Fonte: dados da pesquisa.

ANCINE

Figura 30 – Tipologia SNIIC ANCINE



Fonte: dados da pesquisa.