

Usos participativos das estruturas sociais e linguísticas

Renato Fabbri
renato.fabbri@gmail.com

Oswaldo Novais de Oliveira Junior
chu@ifsc.usp.br



Introdução

Os rastros das atividades em redes sociais permitem a detecção de comunidades e papéis desempenhados pelos indivíduos[1,2]. No entanto, não há uma abordagem focada no participante, em que as medidas e visualizações são voltadas para conscientizá-lo e capacitar seu aproveitamento dos mecanismos sociais[3]. Para isso, destacam-se as áreas de Redes Complexas, Processamento de Linguagem Natural, Dados Linkados e Participação Social. A primeira é dedicada às relações em rede, a segunda à extração de informação dos textos, a terceira aos metadados e ontologias e a última à ação da sociedade civil nos assuntos públicos. Este arcabouço sustenta a apreensão e modelagem dos processos sociais. Como resultado, há contribuições para formalismos nas dinâmicas sociais, implementações computacionais voltadas para o participante e a escrita de textos acadêmicos[1,2,3]. Este trabalho conflui com uma série de trabalhos realizados por acadêmicos (e.g. IFSC/USP, CCNH/UFABC, Lappis/UnB), por órgãos do governos (SG/PR) e pela sociedade civil (PNUD/ONU), em uma iniciativa conjunta estado-sociedade para um portal federal de participação social (Participa.br).

Materiais

Arquivos públicos de mailing lists (GMANE). Dados de grupos e perfis do Facebook. Tweets sobre hashtags e palavras em específico. Nos casos de desenvolvimentos tecnológicos, como ontologias, streamers e visualizações, os materiais foram as APIs, protocolos e conceitualizações.

Métodos

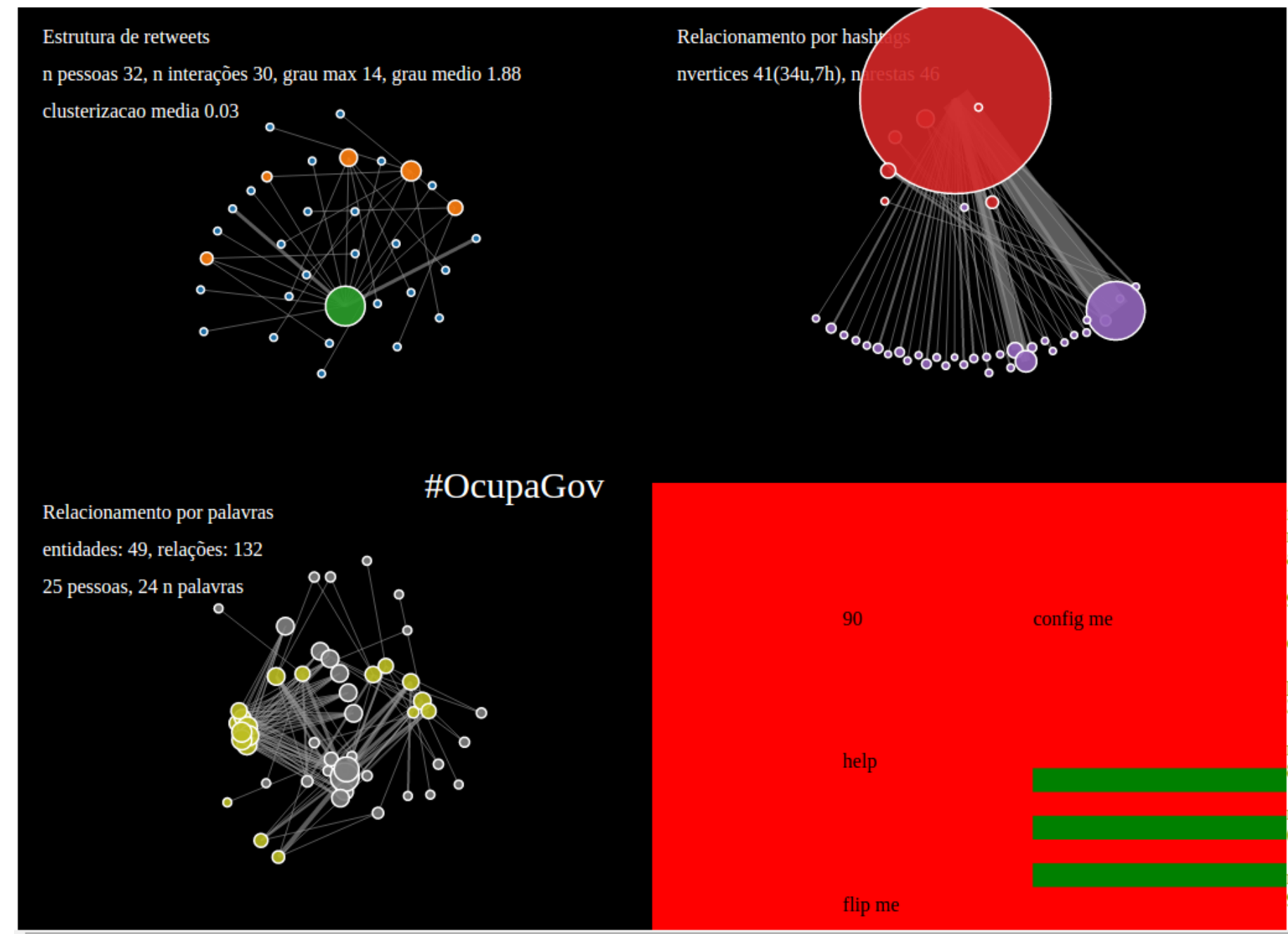
Para os resultados em redes complexas, foi usada a habitual comparação com a Erdős-Renyi com o mesmo número de vértices e arestas[1]. A **estabilidade** foi observada com uma janela deslizante de tamanho fixo de mensagens[2]. A **diferenciação** da produção textual dos setores conectivos (de hubs, intermediários e periféricos) foi observada através da comparação dos histogramas dos usos de recursos linguísticos nas mensagens, com uma adaptação do método de Kolmogorov-Smirnoff[3]. Para as frentes computacionais, são usadas principalmente as linguagens Python e Javascript para as tarefas envolvidas: numéricas, visualizações, enriquecimento semântico, conexão com APIs. Foi desenvolvido um método computacional de levantamento de ontologia orientada aos dados. Um método característico deste trabalho é o diálogo com a comunidades online interessadas ou relacionadas ao pesquisador, no qual o próprio pesquisador é cobaia, o que foi por vezes chamado de "física etnográfica" ou "física antropológica" por parceiros. Todo código e documentação produzida é disponibilizada abertamente, geralmente em repositórios git.

Resultados

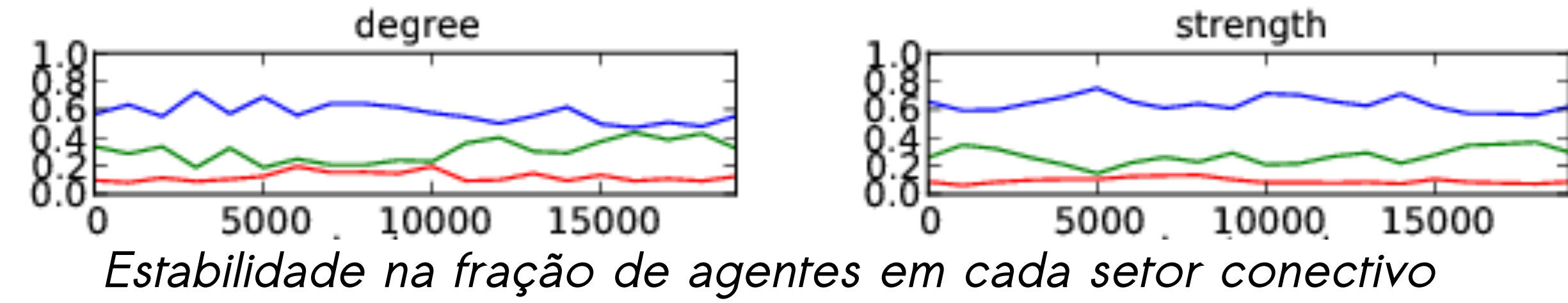
- Física:**
- *) Observada estabilidade nas redes humanas de interação.
 - *) Observada diferenciação no uso dos recursos linguísticos quando a conectividade do participante muda.
 - *) Experimentos de física etnográfica/antropológica, com hipóteses verificadas.
 - *) Resultados não sistematizados, como traços homogêneos em redes livres de escala.

- Computacional:**
- *) Implementações de streaming de estruturas sociais e recursos em nuvem.
 - *) Avanços em dados linkados abertos, com novas ontologias, roteiros de triplificação de bancos relacionais, endpoint SparQL e conexão de legados públicos à LOD (Linked Open Data).
 - *) Novo método de inferência de ontologia orientada a dados.

- Usos:**
- *) Rotinas de coleta e difusão de informação.
 - *) Portal federal de participação social, Participa.br
 - *) Chancela PNUD/ONU para entregar estes conhecimentos e tecnologias à população.
 - *) Telões de monitoramento em tempo real de redes sociais.
 - *) Pesquisa: vários casos e conjunto de dados interessantes. Transbordam implementações, propriedades, hipóteses. São 6 artigos já prontos para enviar, vários outros rascunhados. As publicações e participação em eventos estão nas conclusões.



Telões de streaming de estruturas sociais para acompanhamento de processos midiáticos.

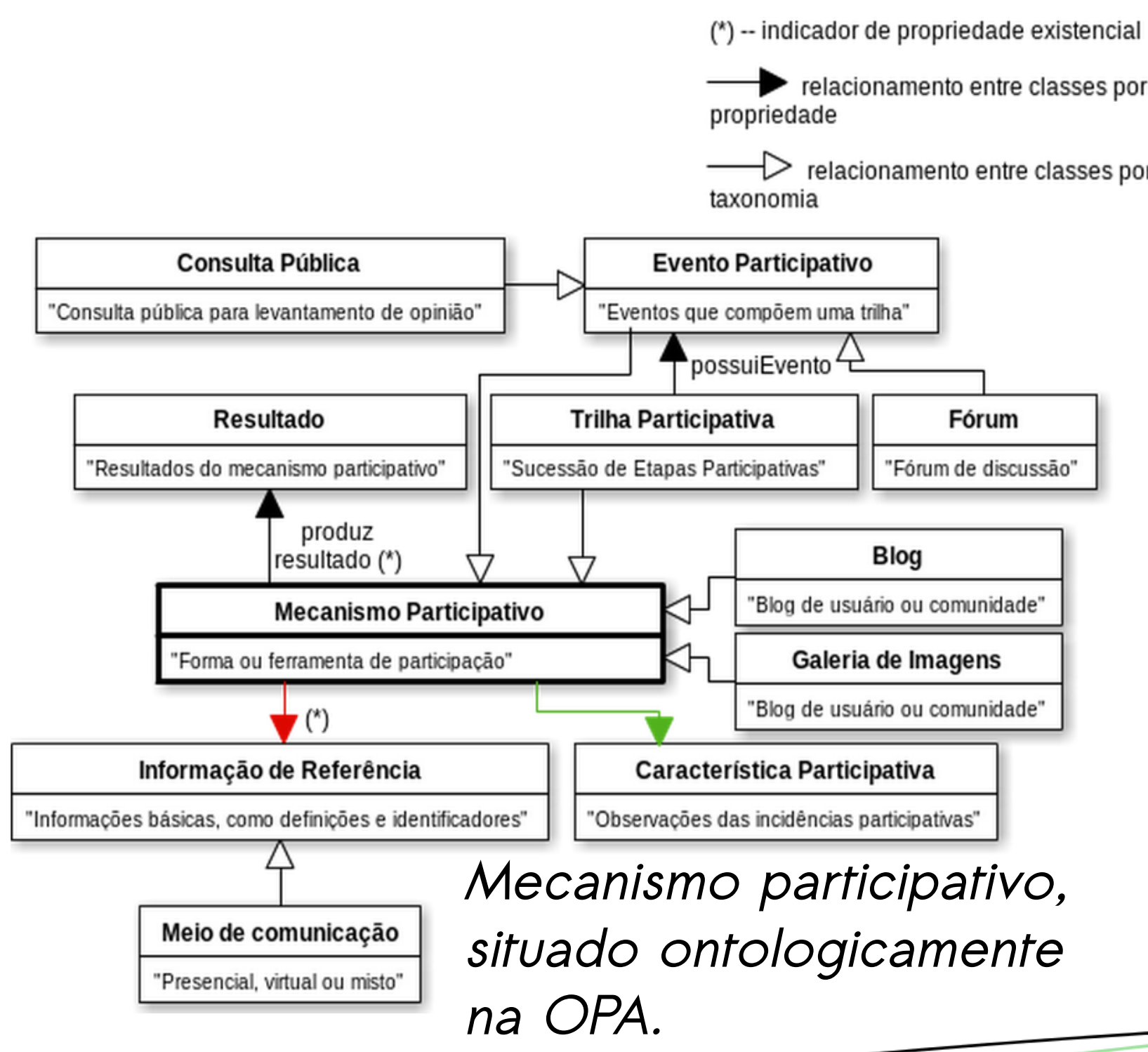


list\measure	H-P	H-I	I-P
CPP	7.01	4.89	7.95
LAD	9.82	6.13	8.58
LAU	9.66	5.44	7.45
ELE	5.78	2.84	4.69

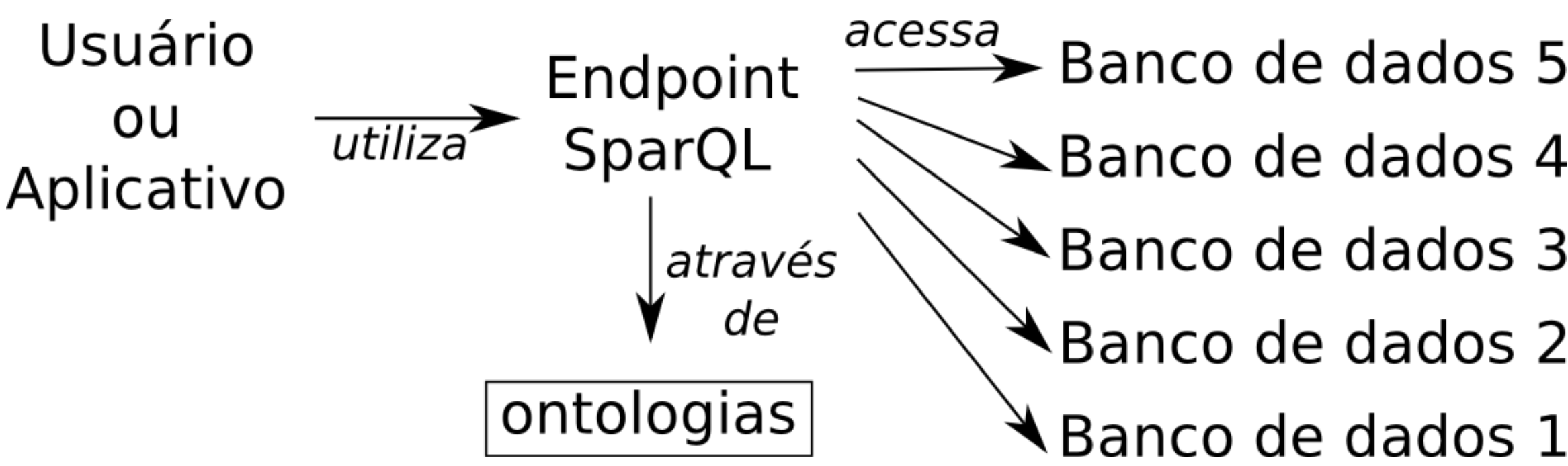
Diferenciação da produção textual dos setores Hub, Intermediário e Periférico.

ranking	Participa	AA
1	participação => 717	git => 2052
2	social => 548	commit => 1985
3	sociedade => 526	fazer => 1612
4	governo => 367	ainda => 1171
5	forma => 337	agora => 1066
6	políticas => 327	tentando => 981
7	públicas => 310	aa => 978
8	brasil => 302	vou => 920
9	nacional => 266	lendo => 835
10	todos => 258	email => 789
11	educação => 251	artigo => 709
12	política => 249	scilab => 568
13	deve => 232	ver => 566
14	civil => 231	terminando => 461
15	proposta => 225	wiki => 419
16	acesso => 224	dando => 388
17	federal => 223	pd => 388
18	país => 222	sip => 374
19	direito => 218	testando => 348
20	pode => 218	pouco => 345

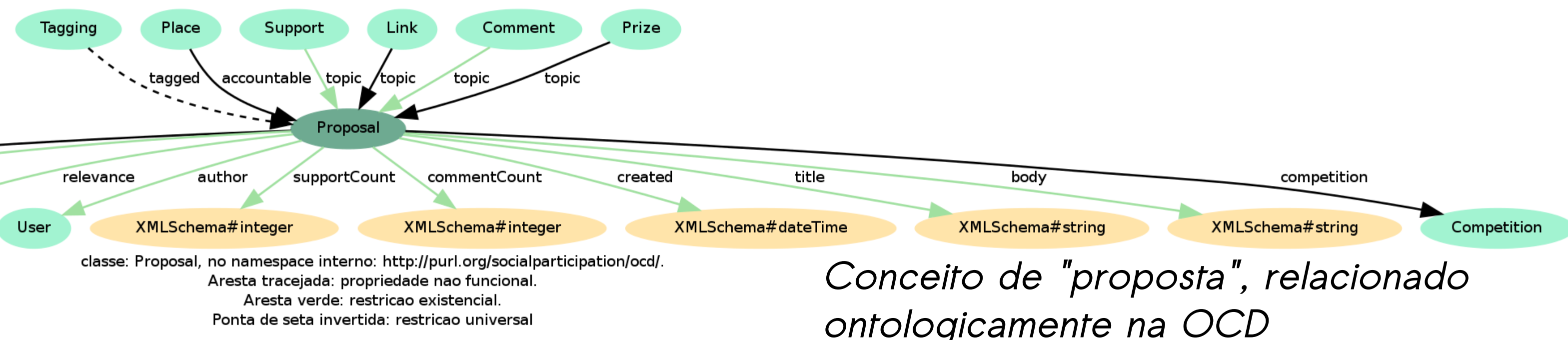
Palavras mais incidentes em instâncias diferentes



Mecanismo participativo, situado ontologicamente na OPA.



Esquema comum de uso de um endpoint SparQL para acessar dados semi-estruturados (em RDF) de fontes diversas



Conceito de "proposta", relacionado ontologicamente na OCD

Conclusões e Trabalhos Futuros

Embora esta pesquisa tenha sido acolhida por parceiros e iniciativas diversas, os aspectos práticos e multidisciplinares tornam grande o tempo de dedicação necessário. O ideal seria realizar a pesquisa com uma equipe de programadores, linguistas, comunicadores e físicos. Foram publicados artigos sobre esta pesquisa e sobre pesquisas menos centrais para o doutorado. Houve bastante participação em eventos, alguns acadêmicos, como o fechamento do Congresso Internacional de Teoria Crítica (ROMA). Os feitos acumulam o suficiente para caminhar na direção de uma qualificação. A publicação dos artigos já escritos, e dos rascunhados, e o posterior aprofundamento com parceiros e literatura, deve acompanhar as investidas finais deste doutorado. Para finalização da parte tecnológica, está sendo desenvolvido um sistema de navegação pelos dados interligados, enriquecido através de análises de instâncias e coletivos, recomendação de recursos a partir de um recurso de referência, explicitação dos possíveis usos dos recursos e métodos de recomendação, e entrega dos métodos como código computacional aberto.

Referências

(1) Easley, David, and Jon Kleinberg. Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press, 2010.

(2) Allemang, Dean, and James Hendler. Semantic web for the working ontologist: effective modeling in RDFS and OWL. Elsevier, 2011.

