

A complex network diagram with numerous nodes of varying sizes and colors (grey, black, white) connected by solid and dotted lines, set against a background of scattered grey dots.

# Barômetro da tecnociência

---

Techno-scientific Issues in the  
Public Sphere (TIPS)

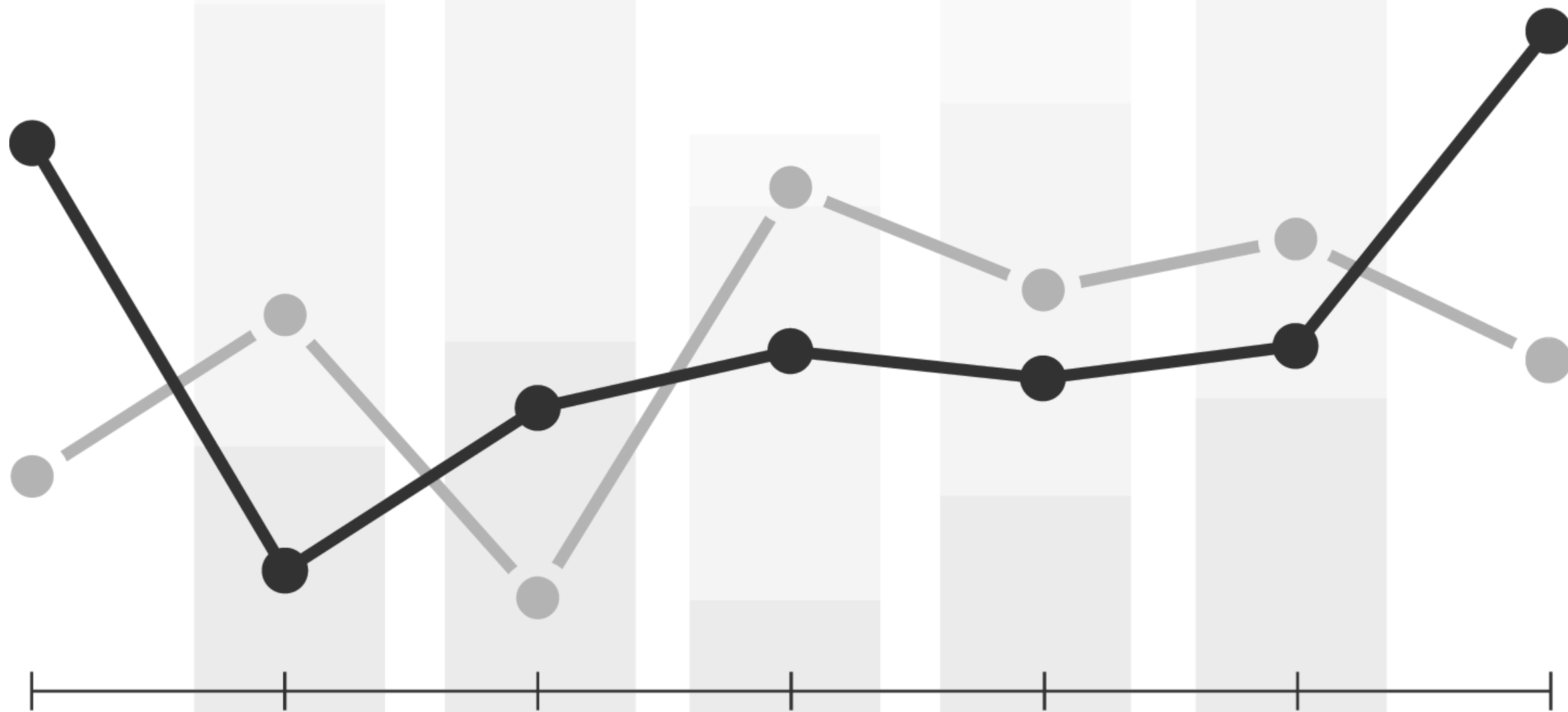
---

Maio de 2018  
Lucas V. Emerick  
Rodrigues



# Introdução

- Uma ferramenta para avaliação da circulação de conteúdo científico na esfera pública
- O que é científico? O que é esfera pública?
- A Ciência para o Barômetro



## **Sobre a Tecnociência**

Os processos e produções científicas são atividades exercidas de forma coletiva, e com isso queremos dizer que são também:

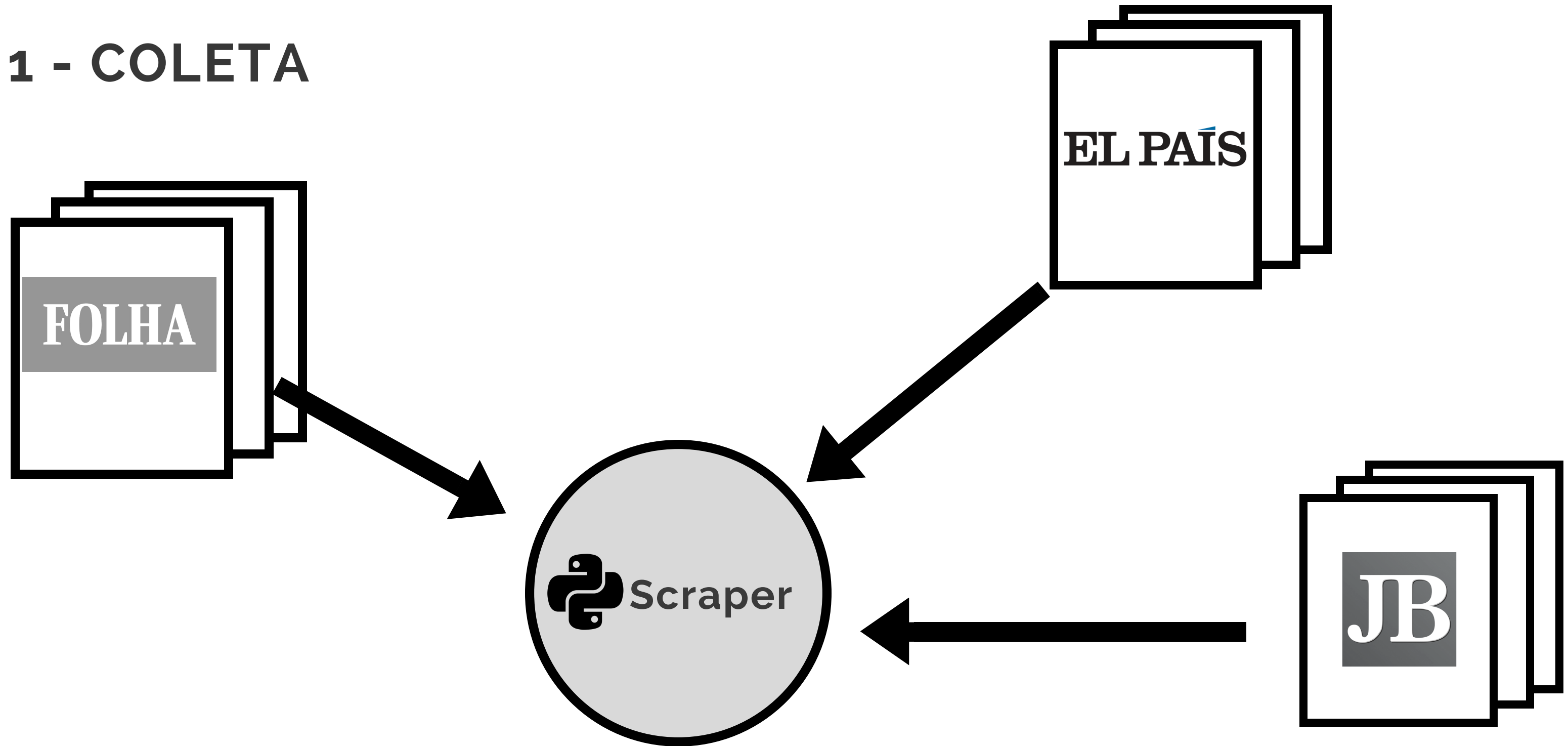
- Feitas por alguém;
- Feitas dentro de uma instituição;
- Feitas dentro de um amplo processo de comunicação;
- Feitas através de desenhos (design), testes e desenvolvimento usando instrumentos;
- Feitas por meio de objetos específicos e enfatizando artefatos específicos;
- Descrita por narrativas que usam palavras e expressões específicas;

# O que é a ciência para o barômetro

- Menção a um cientista ou engenheiro
- Menção a um centro de pesquisa
- Menção a um jornal científico
- Menção a uma disciplina científica
- Referência a algum processo de pesquisa e/ou inovação tecnológica
- Menção a uma descoberta, uma inovação, um instrumento científico ou aparato médico.

# Metodologia

## FASE 1 - COLETA

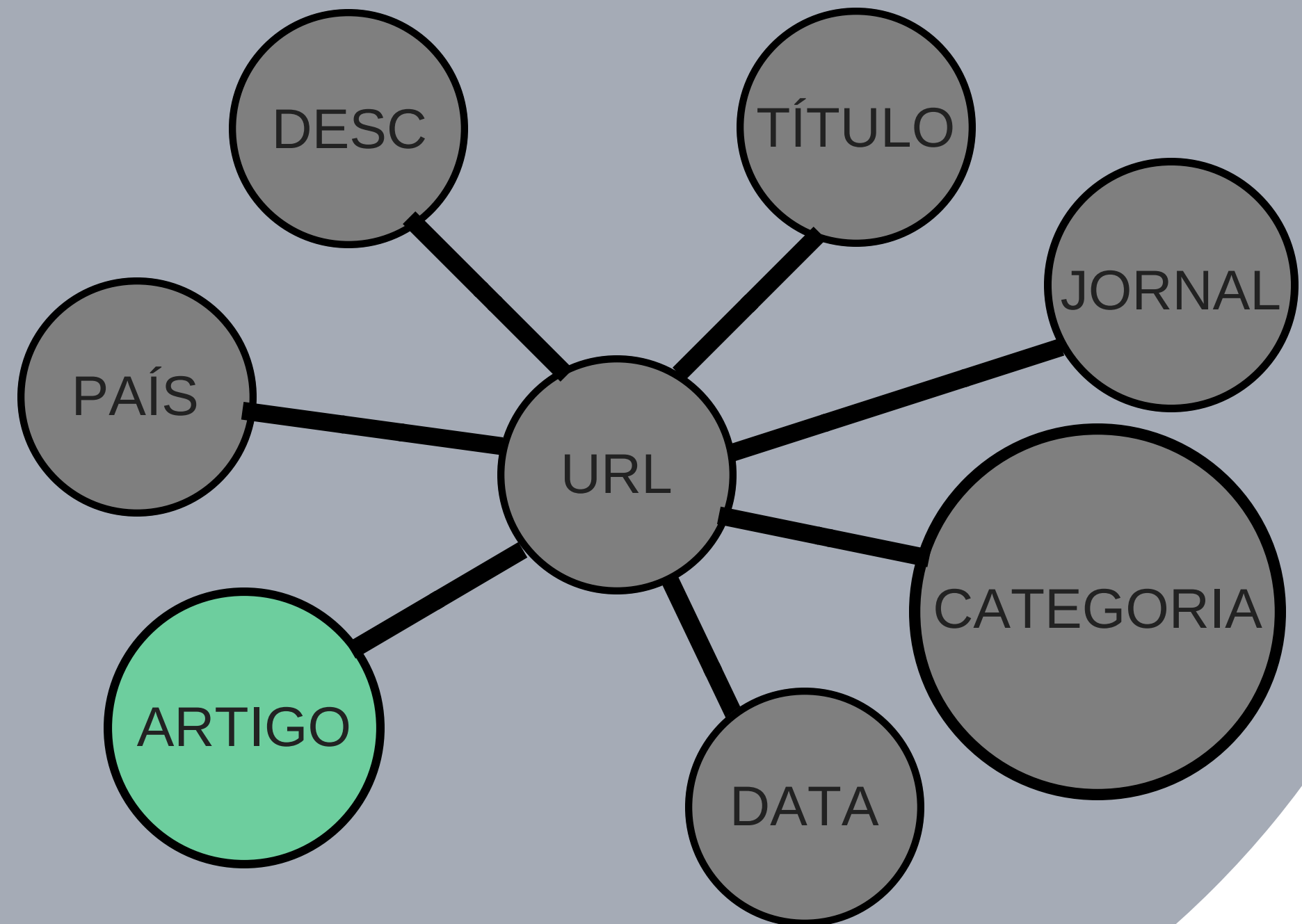


## FASE 1 - COLETA

Foi criada uma rotina que diariamente coleta os artigos publicados em vários canais de notícia da América Latina, de Portugal e da Espanha (Iberoamérica).

O texto dos artigos é coletado assim como o título, a descrição, a data, a seção, o país e o jornal. Esse agregado de informação é enviado a um **banco de dados de grafo** que relaciona e armazena cada artigo dos canais

## FASE 2 - ARMAZENAMENTO



## FASE 2 - ARMAZENAMENTO

Os items são inseridos no banco de dados Neo4j em forma de nós e arestas. Os artigos e suas informações são representados em formas de nós e a relação existente entre eles são as arestas. Por exemplo, a categoria 'Esportes' categoriza um certo artigo. Numa notação de grafo esta relação estaria disposta da seguinte forma:

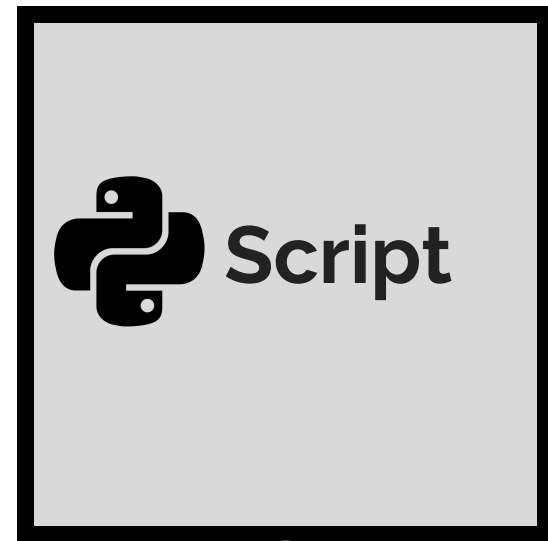
**(Esportes) - [Categoriza]->(Artigo)**

Um nó representando a categoria 'Esportes', outro nó representando um artigo e 'Categoriza' representando a relação que o nó 'Esportes' exerce sobre o artigo.

Os nós se mesclam quando os nomes são iguais, então na fase do armazenamento várias aglomerações de artigos já são estabelecidas em torno de suas categorias (as seções), dos canais e dos países em que foram publicados

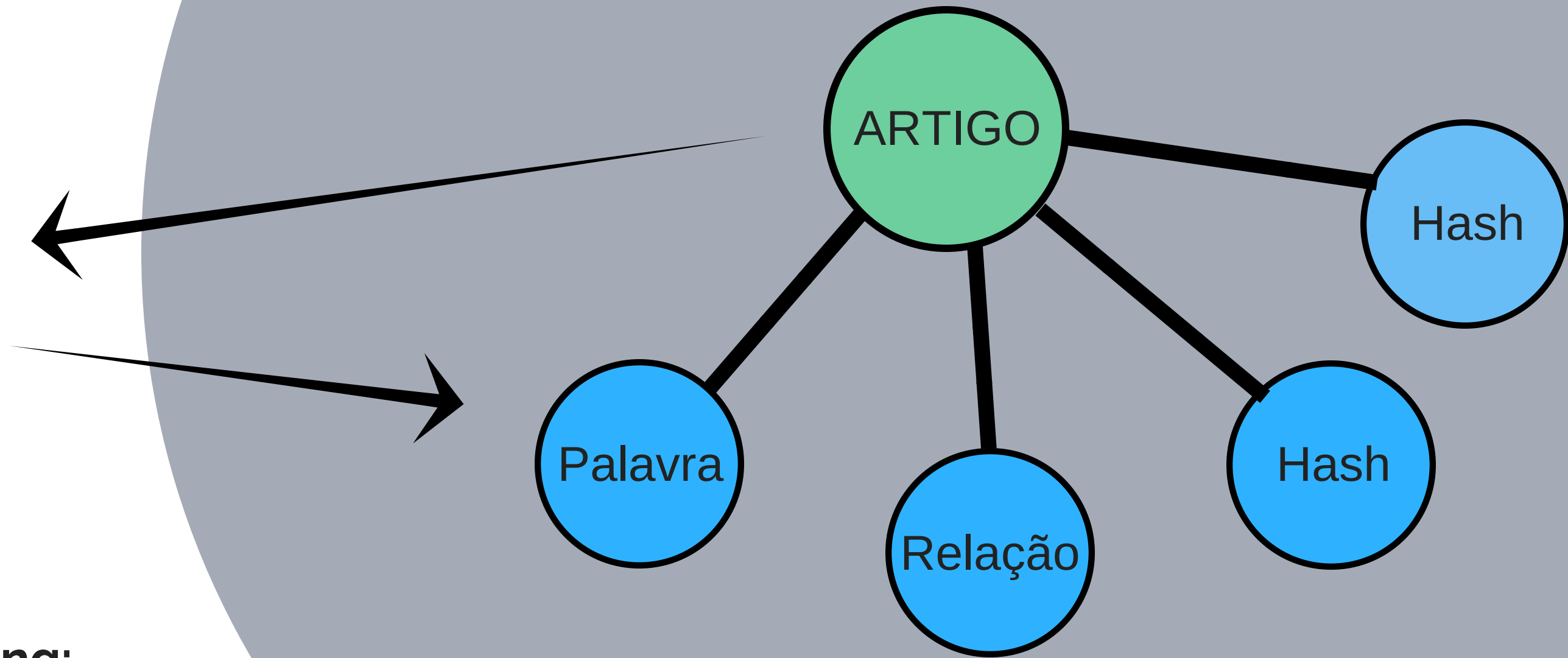


# FASE 3 - PROCESSAMENTO



Knowledge Engineering:

- Naive Bayes
- Support Vector Machine



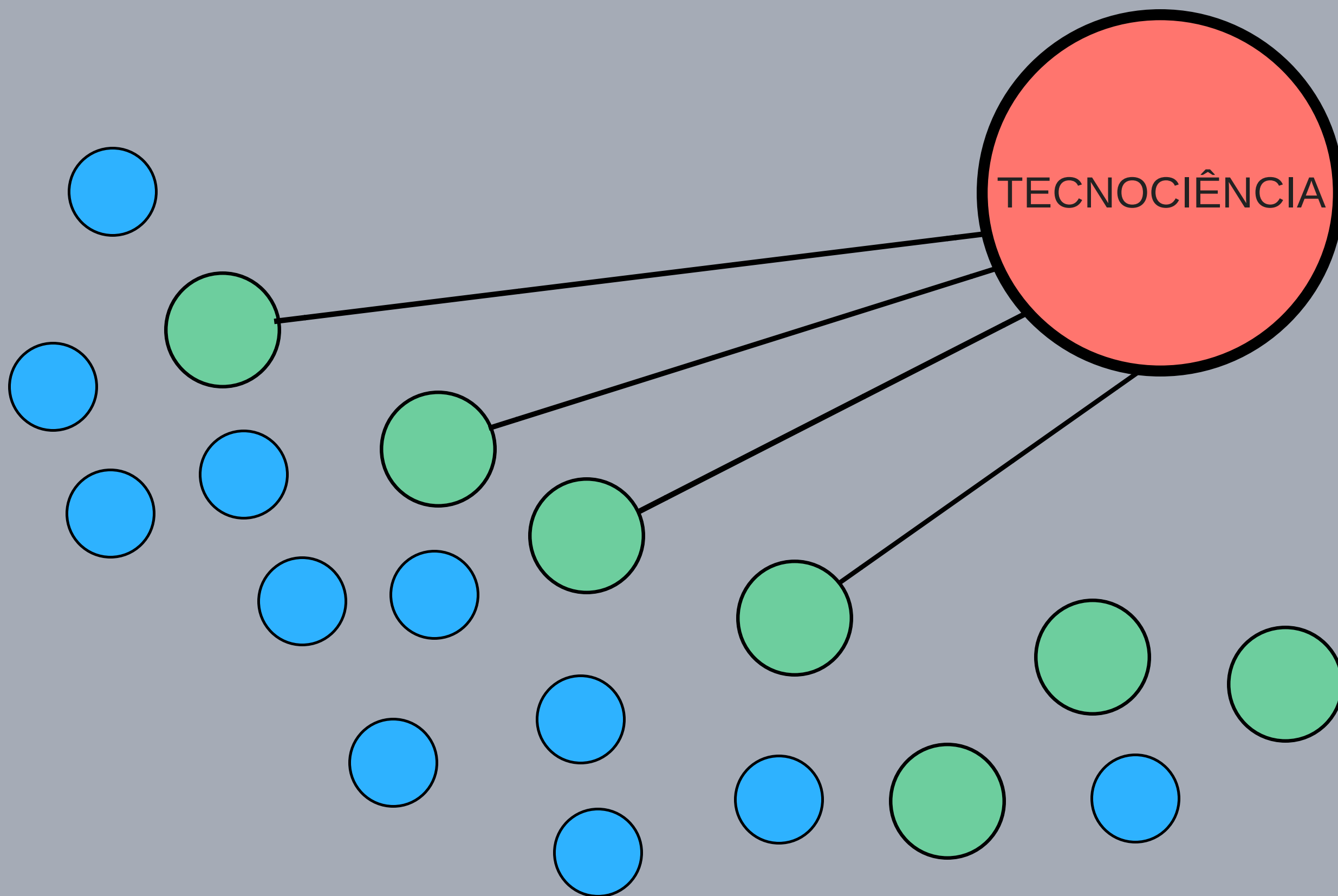
## **FASE 3 - PROCESSAMENTO**

Para um aproveitamento eficaz dos artigos, seus textos são sujeitos a alguns algoritmos que os fragmentam em unidades e relações significativas, ou seja, criam uma rede semântica do texto, relacionando das mais diversas formas os substantivos, verbos e adjetivos presentes nele.

Essas palavras que surgem do processamento são relacionadas ao texto de origem, sendo também representadas e armazenadas na forma de nós que relacionam-se entre si na forma de arestas.

Já nessa fase do processo os nós de palavras iguais se mesclam produzindo assim aproximações entre artigos pela correlação de palavras.

# FASE 4 - CRIAÇÃO DE UM CORPUS DE CONHECIMENTO

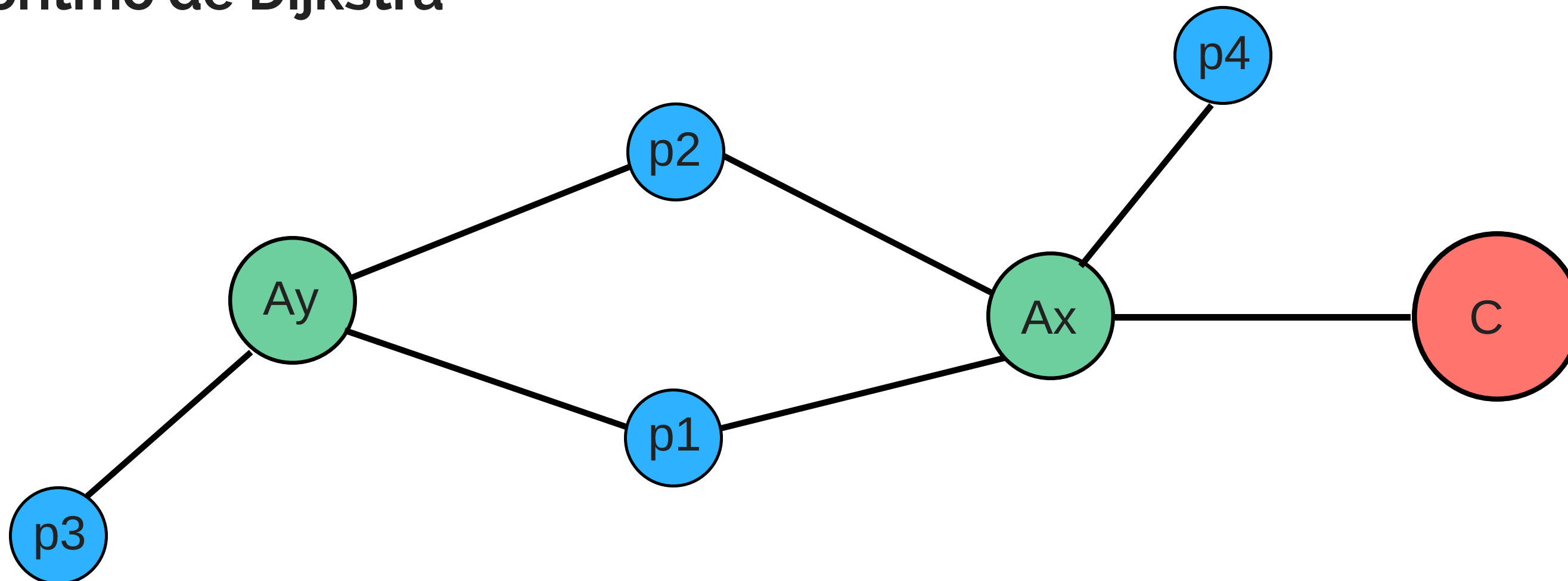


## **FASE 4 - CRIAÇÃO DE UM CORPUS DE CONHECIMENTO**

- O Corpus de Conhecimento é construído a partir da seleção de um certo número de artigos que x pesquisadorx tomar como relativo a essa classe, no caso do projeto: A Ciência.
- Pode-se agregar artigos que tenham relação com Tecnociência, assim como Sociologia, ou Agricultura, ou até OVNI's.
- Essa curadoria/seleção de artigos é necessária para que existam artigos de referência, ou seja artigos que servirão como parâmetro para classificação de outros artigos.
- O Corpus é também apenas mais um nó que será conectado aos nós dos artigos selecionados como relacionados a classe.

# FASE 5 - CLASSIFICAÇÃO

## Algoritmo de Dijkstra



Quantos são os menores caminhos entre C e Ay?

=

O quanto esse Artigo está relacionado ao Corpus?

## FASE 5 - CLASSIFICAÇÃO

Depois do Corpus da ciência definido e relacionado aos artigos que se adequam a classe, restam os artigos que não estão diretamente ligados ao Corpus. Esses artigos podem ou não ser científicos e cabe ao classificador fazer essa distinção.

A partir das aproximações estabelecidas entre os artigos pela correlação de palavras é possível medir o quanto um artigo não diretamente relacionado a uma classe está relacionado a ela. Por meio do algoritmo de Dijkstra torna-se viável descobrir e calcular todas as intersecções entre um artigo e uma classe, e a partir desse número resultante é possível definir um ponto de corte que enquadrará ou não o artigo dentro da classe em questão.