



IME-USP

ALGORITMO MULTINÍVEL DE DETECÇÃO DE COMUNIDADES EM REDES COMPLEXAS

STEVEN KOITI TSUKAMOTO

Orientadores: Prof. Dr. Marcelo Finger
Dra. Ana Paula Appel (IBM Research)

Colaborador: Prof. Dr. Estevam Hruschka (UFSCar)

INTRODUÇÃO

O surgimento e crescimento de dados modelados em grafos, como as redes complexas, tornou possível a criação de novas áreas de pesquisa como a mineração de grafos. Nesta área os assuntos que se destacam são a detecção de comunidades, a extração de propriedades estatísticas e a predição de ligações.

A detecção de comunidade pode ser utilizada para o pré-processamento do grafo a fim de quebrá-lo em grupos que podem ter suas propriedades mineradas mais facilmente.

MOTIVAÇÃO

Redes complexas estão repletas de recursos e ao extrair essa grande quantidade de informação chega-se a um ponto em que não conseguimos mais analisar individualmente as instâncias e necessitamos enxergar de uma outra maneira o problema, de modo mais rápido e fácil. Com os métodos de detecção de comunidades podemos analisar mais abrangentemente como essas instâncias se conectam e se comportam dentro do grafo.

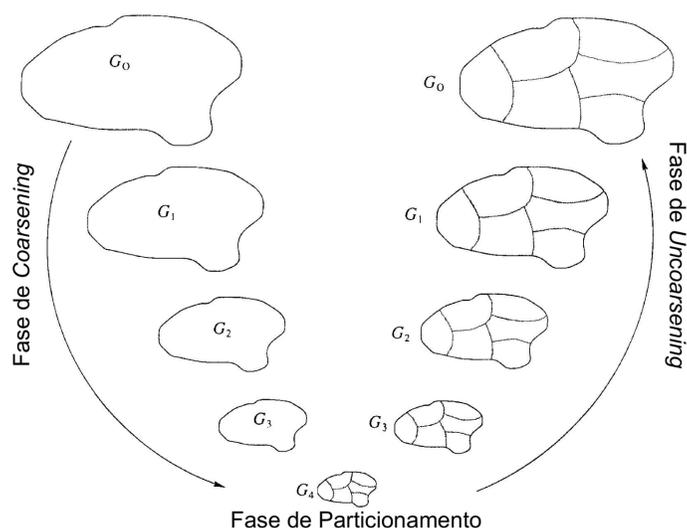
Um dos maiores benefícios seria então poder compreender como o sistema evolui e de que maneira os subgrupos se aglomeram.

OBJETIVOS

O objetivo deste trabalho consiste em encontrar métodos de detecção de comunidades que se adequem à sistemas dinâmicos e escaláveis. Dividimos os objetivos em:

- Implementação de um algoritmo multinível para detecção de comunidades.
- Análise comparativa dos resultados aplicados a uma rede complexa.

ALGORITMO MULTINÍVEL



Fase de coarsening: o grafo é transformado sucessivamente em grafos menores através da contração de suas arestas.

Fase de particionamento: o grafo reduzido é particionado em subgrupos.

Fase de uncoarsening: os particionamentos são projetados para cada um dos grafos intermediários gerados na fase de coarsening até que se retorne ao grafo original.

SISTEMA NELL

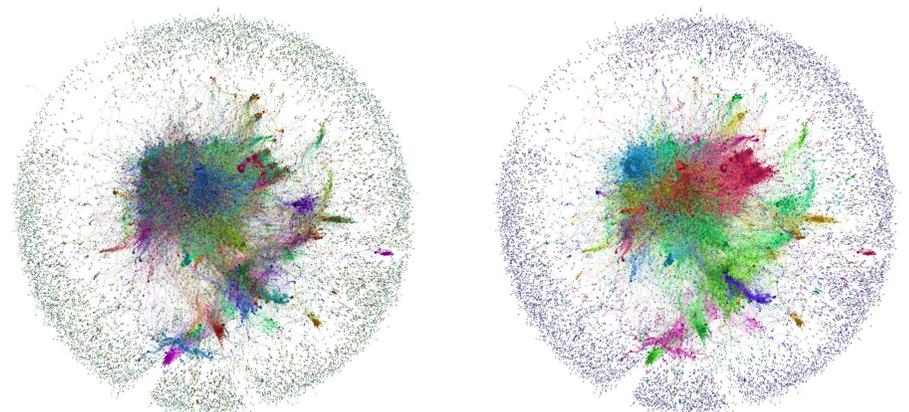
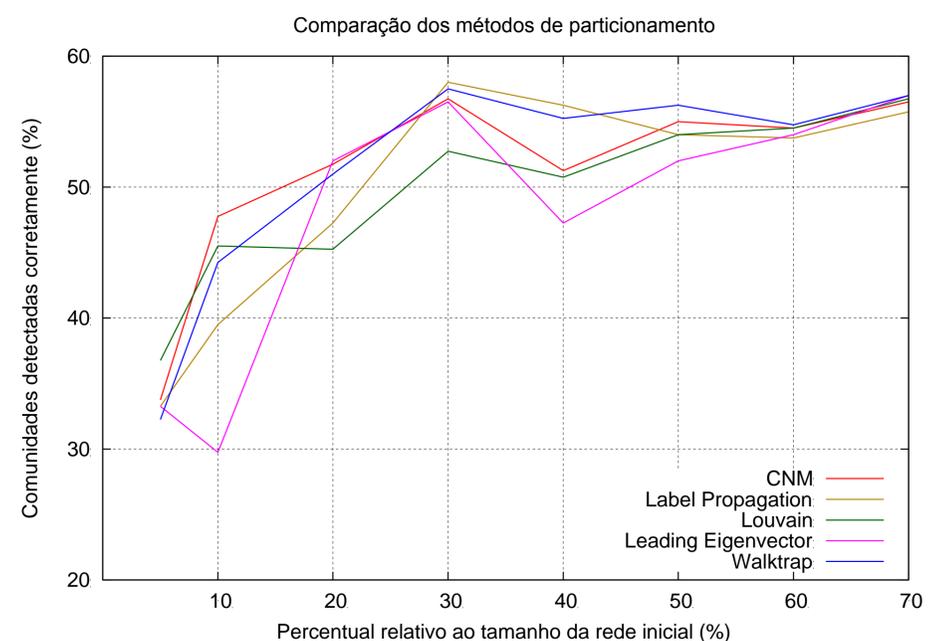
O Principal estudo de caso deste trabalho é o sistema NELL (Never-Ending Language Learner). Desenvolvido por pesquisadores da Universidade Carnegie Mellon e da Universidade Federal de São Carlos, o NELL é um sistema de aprendizado contínuo que tem como objetivo percorrer a internet e extrair conhecimento de milhares de textos existentes na rede. As duas principais tarefas executadas são:

- **Leitura:** Extrair dados encontrados em textos de páginas web e popular uma base de conhecimento com fatos estruturados.
- **Aprendizado:** a cada dia conseguir melhorar sua capacidade de leitura, extraindo fatos cada vez mais precisos e exatos.

Todo o conhecimento adquirido pode ser modelado na forma de um grafo, onde cada instância possui uma categoria atrelada.

RESULTADOS

O Grafo obtido da base de conhecimento do NELL possui um total de 69.837 vértices e 134.700 arestas.



À esquerda, imagem do grafo colorido pelas categorias. À direita, utilizando algoritmo multinível de detecção de comunidades.

CONSIDERAÇÕES FINAIS

Grandes reduções do grafo na fase de coarsening mostraram resultar em um baixo índice de acerto na detecção de comunidades de mesma categoria. Também podemos notar que em certo ponto, todos os métodos de particionamento convergem para uma mesma taxa de comunidades corretamente detectadas.

Algumas combinações da estratégia multinível com algoritmos de particionamento apresentaram boas soluções de detecção, obtendo taxas razoáveis de acerto.



Read the Web
Research Project
Carnegie Mellon University