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)

Redes Complexas

"A **Complex Network** is a graph (network) with non-trivial topological feature – features that do not occur in simple networks but often occur in graphs modelling **real systems**."

Exemplos: Redes Sociais, Web, Internet, Cadeias Alimentares, Redes de Energia Elétrica

NELL

NELL (Never-Ending Language Learner)

NELL

NELL (Never-Ending Language Learner)

1. **Leitura** – Extrair fatos de textos encontrados em milhões de páginas web.
2. **Aprendizado** – Melhorar sua competência de leitura, extraindo fatos cada vez mais precisos.

NELL

NELL (Never-Ending Language Learner)

Base de conhecimento inicial:

- Categorias – ex: artistas, jogadores, livros, cidades
- Relações – ex: jogaNoTime (atleta, timeEsportivo)

Motivações

1. Analisar o crescimento da rede

Motivações

1. Analisar o crescimento da rede
2. Instâncias categorizadas incorretamente

Motivações

1. Analisar o crescimento da rede
2. Instâncias categorizadas incorretamente
3. Predição de links

Objetivos

- Aplicar métodos de detecção de comunidades
- Fazer uma análise dos resultados obtidos

Grafo do NELL

Grafo do NELL

vértices: 679.292

arestas: 441.329

componentes conexas: 310.983

Tamanho da maior componente: 51.233

Detecção de comunidades

- Girvan-Newman

Detecção de comunidades

- Girvan-Newman
- CNM (Clauset-Newman-Moore)

Detecção de comunidades

- Girvan-Newman
- CNM (Clauset-Newman-Moore)

comunidades: 312.916

99,25% das comunidades de tamanho 2

Detecção de comunidades

- Girvan-Newman
- CNM (Clauset-Newman-Moore)

comunidades: 312.916

99,25% das comunidades de tamanho 2

Ex:

country:brazil -> url:http://en.wikipedia.org/wiki/Brazil

product:macbook -> url:http://en.wikipedia.org/wiki/Macbook

Soluções

Removendo instâncias relacionadas à categoria url

- aproximadamente 90% grafo

Soluções

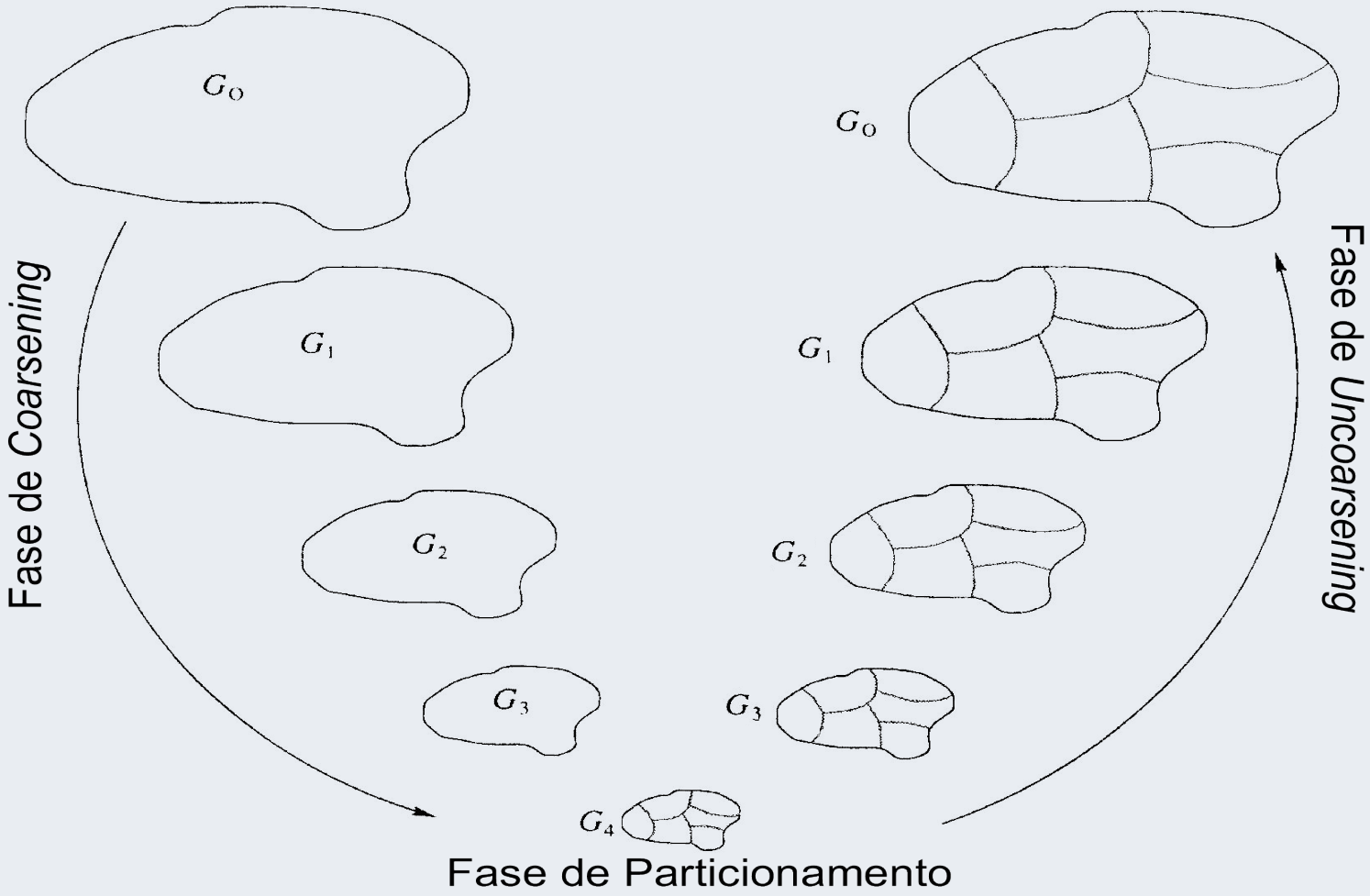
Removendo instâncias relacionadas à categoria url

- aproximadamente 90% grafo

Detecção de comunidade

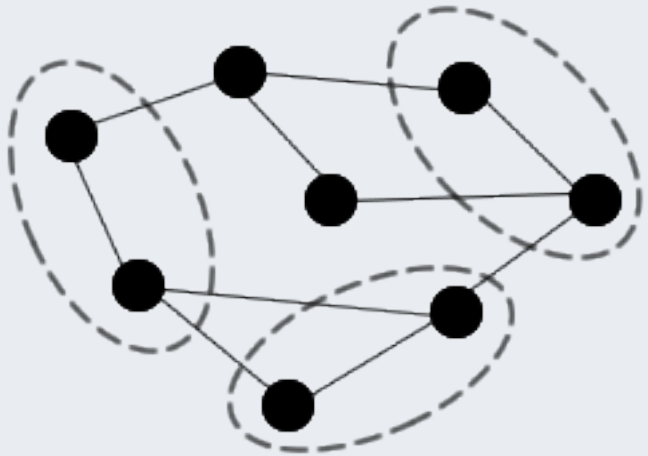
- Metis

Algoritmo Multinível



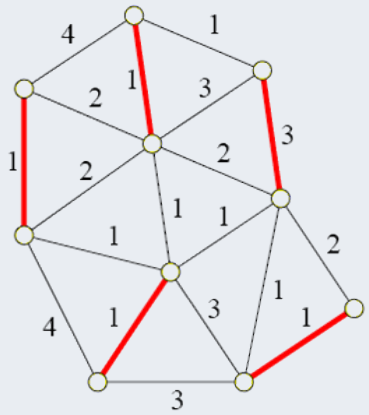
Fase de coarsening

- RM (Random Matching)
- HEM (Heavy-Edge Matching)
- MHEM (Modified Heavy-Edge Matching)
- LEM (Light-Edge Matching)

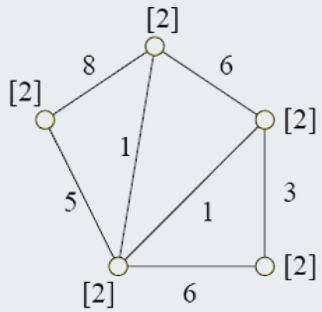


Fase de coarsening

Random Matching

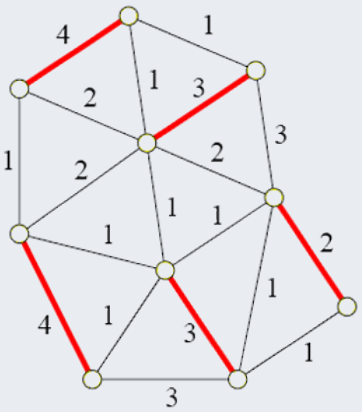


Peso total das arestas = 37

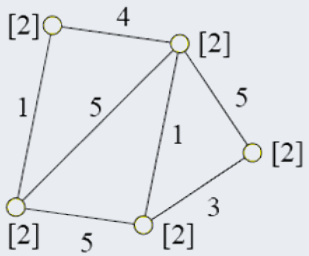


Peso total das arestas = 30

Heavy-edge Matching



Peso total das arestas = 37



Peso total das arestas = 21

Fase de particionamento

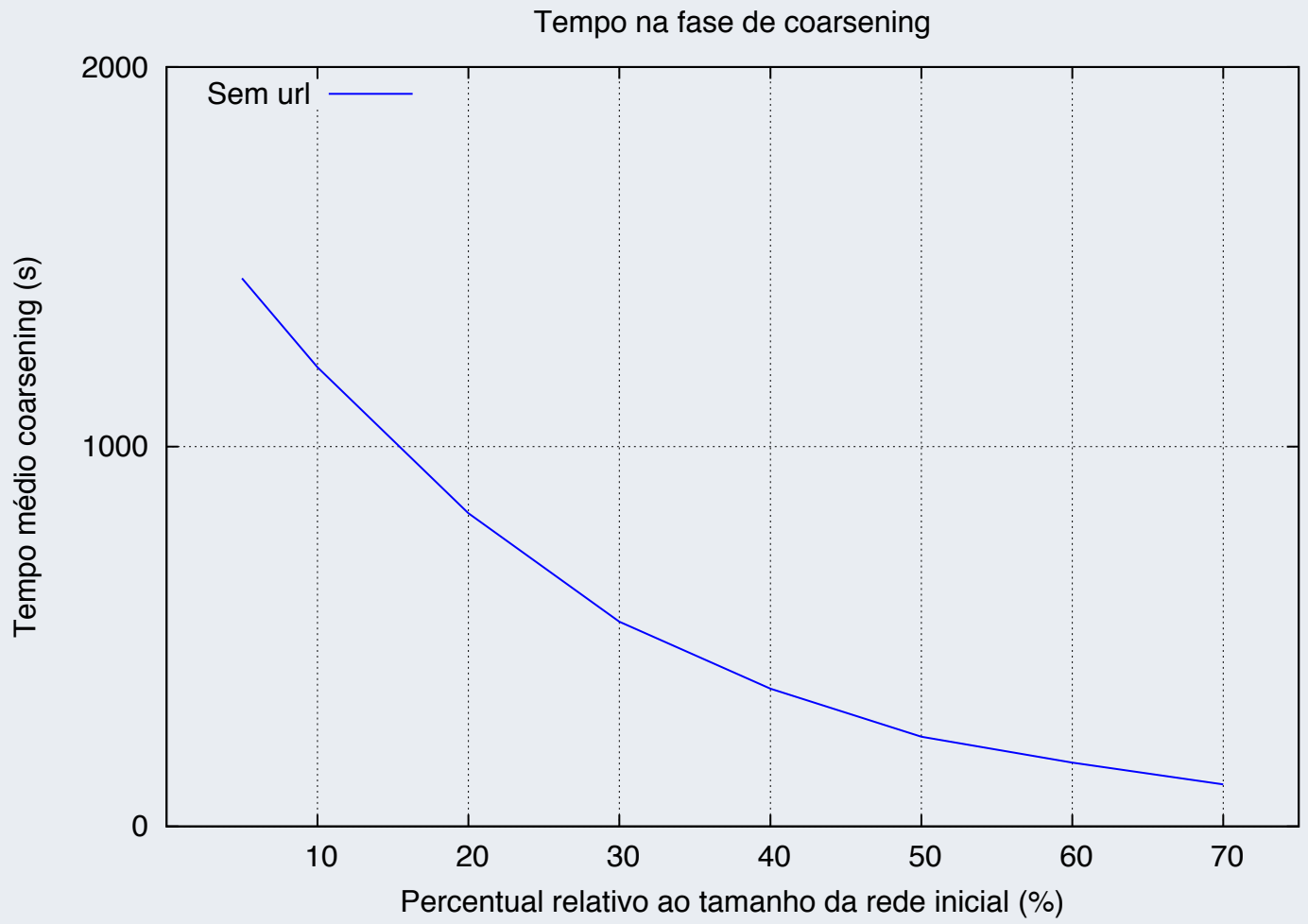
SNAP

- Girvan-Newman
- CNM

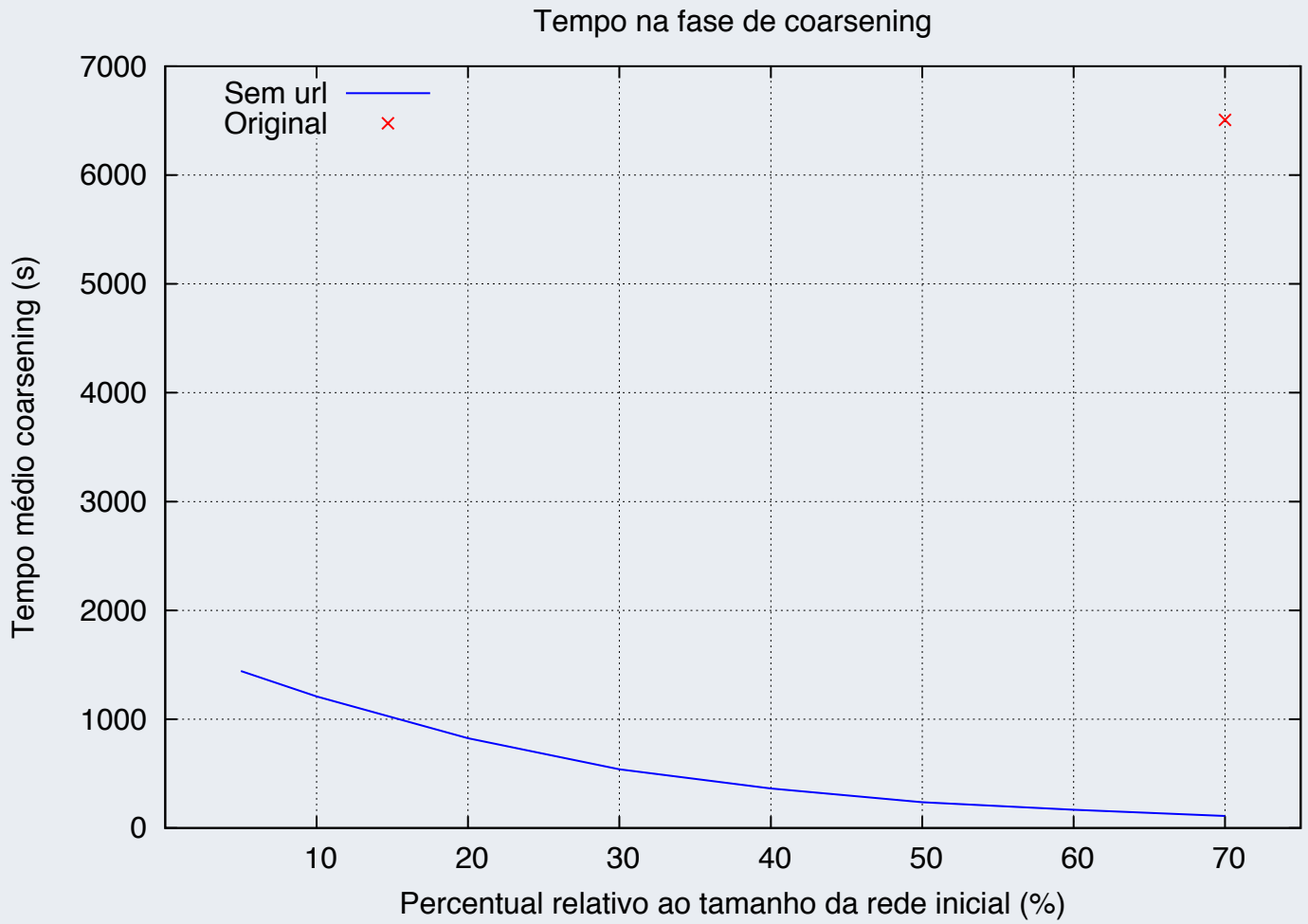
Igraph

- Louvain
- Label propagation
- Leading eigenvector
- Walktrap

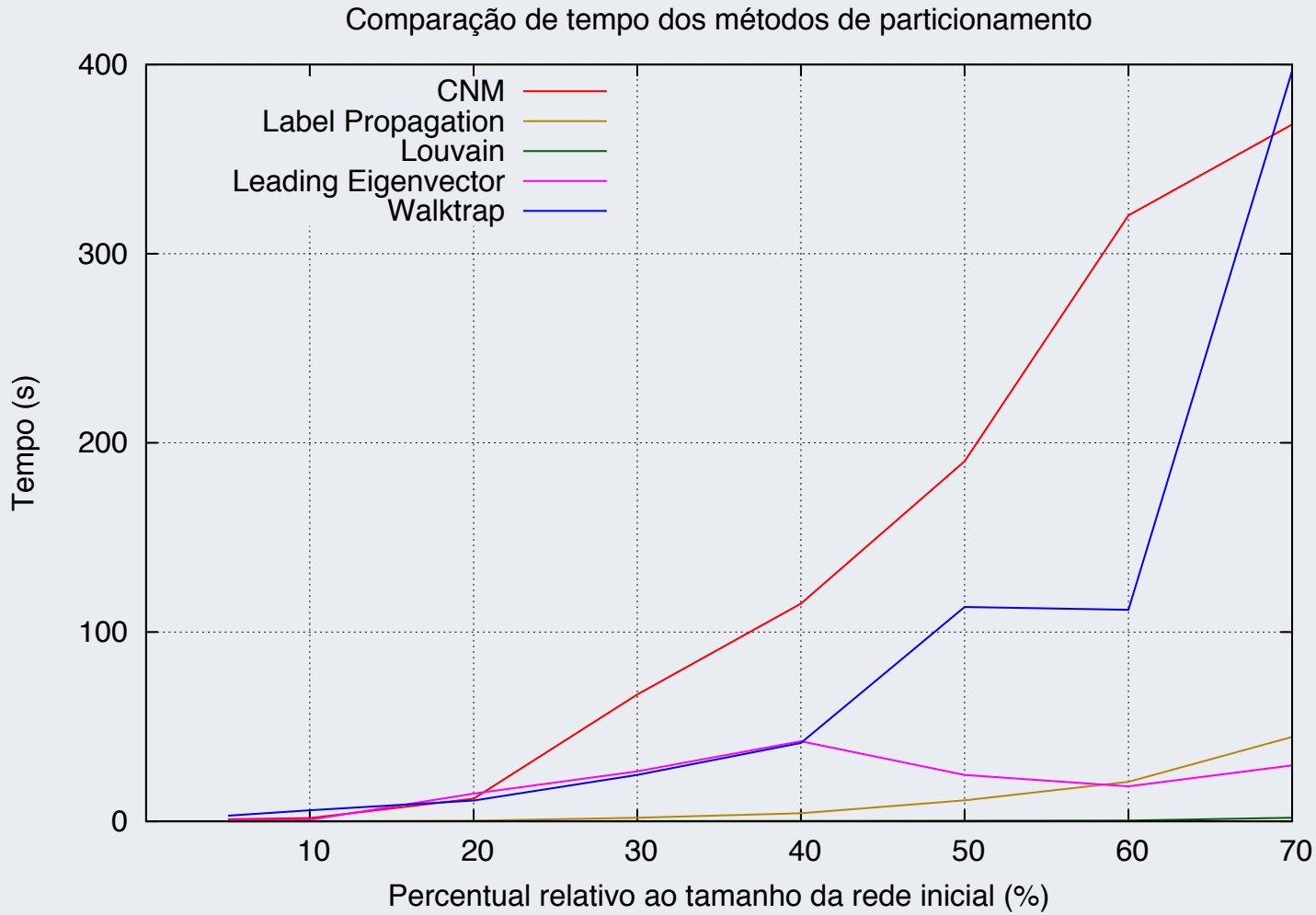
Resultados


Tempo na fase de coarsening
Sem url
2000
1000
0
Tempo médio coarsening (s)
10
20
30
40
50
60
70
Percentual relativo ao tamanho da rede inicial (%)

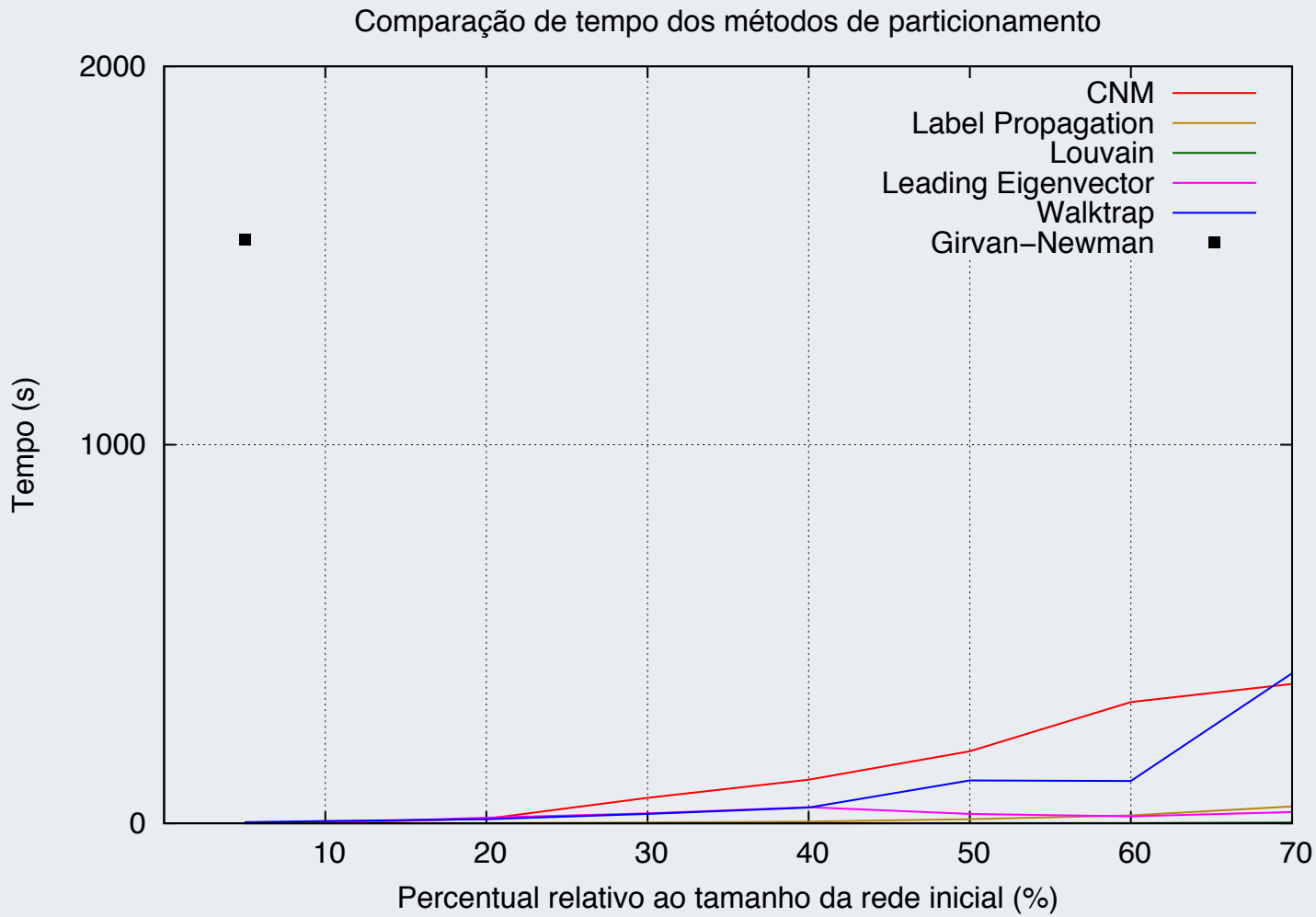
Resultados


Tempo na fase de coarsening
Sem url
Original
Tempo médio coarsening (s)
7000
6000
5000
4000
3000
2000
1000
0
10
20
30
40
50
60
70
Percentual relativo ao tamanho da rede inicial (%)

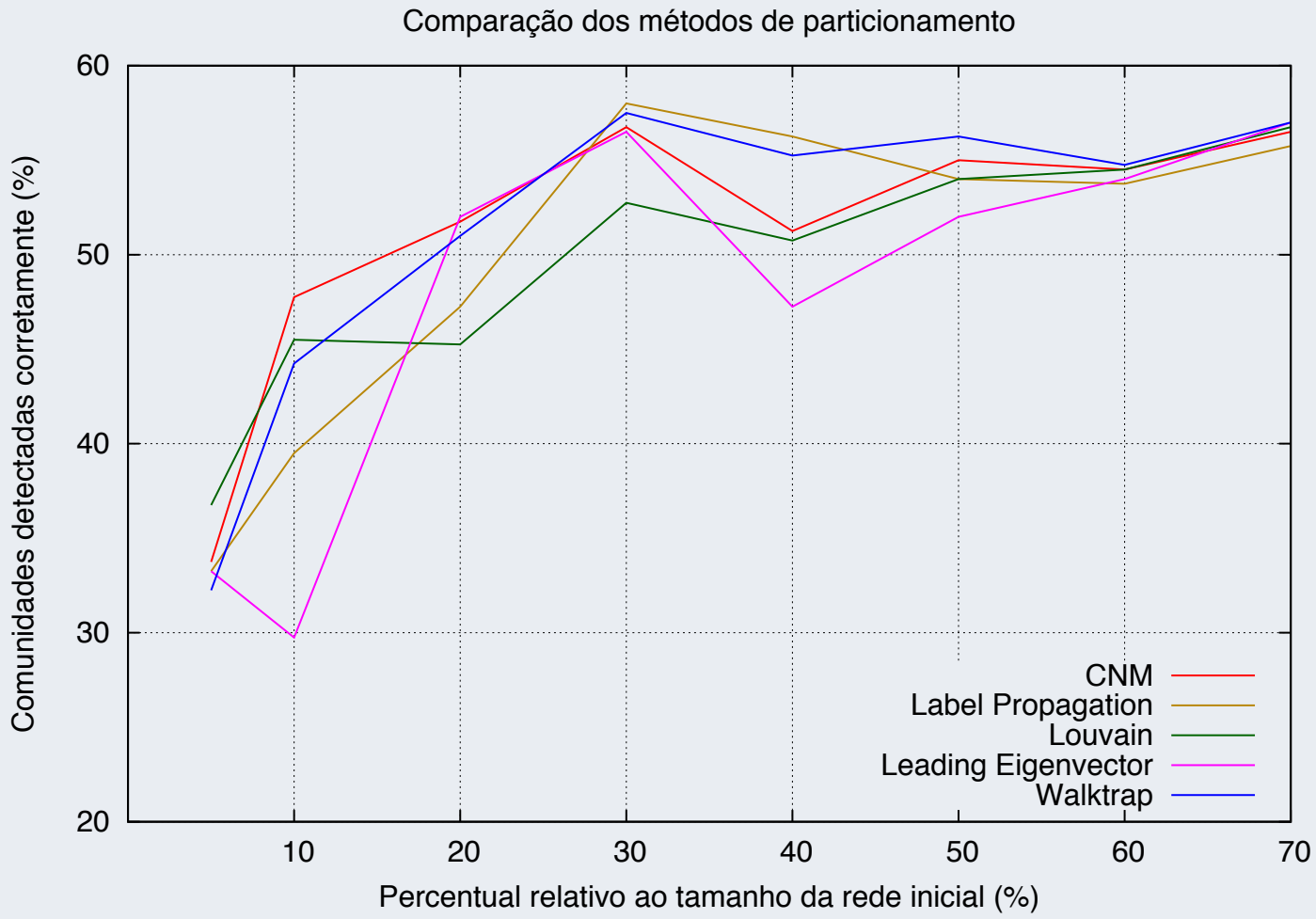
Resultados


Comparação de tempo dos métodos de particionamento
CNM
Label Propagation
Louvain
Leading Eigenvector
Walktrap
Tempo (s)
0
100
200
300
400
10
20
30
40
50
60
70
Percentual relativo ao tamanho da rede inicial (%)

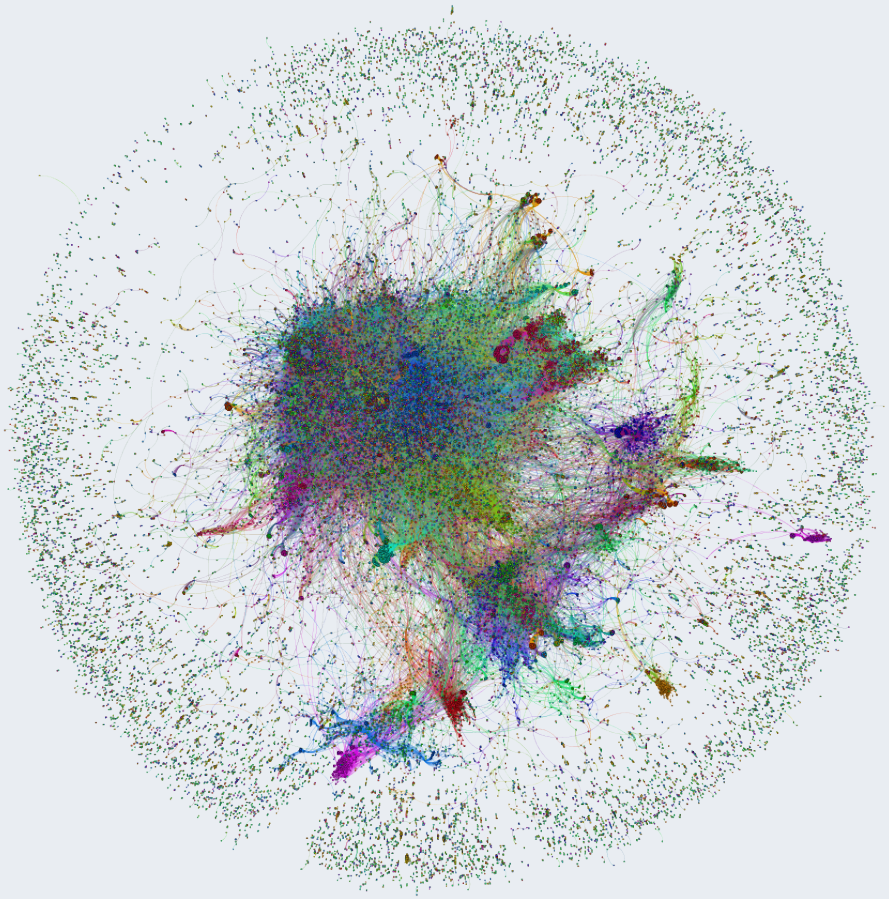
Resultados


Comparação de tempo dos métodos de particionamento
2000
1000
0
Tempo (s)
10
20
30
40
50
60
70
Percentual relativo ao tamanho da rede inicial (%)
CNM
Label Propagation
Louvain
Leading Eigenvector
Walktrap
Girvan–Newman

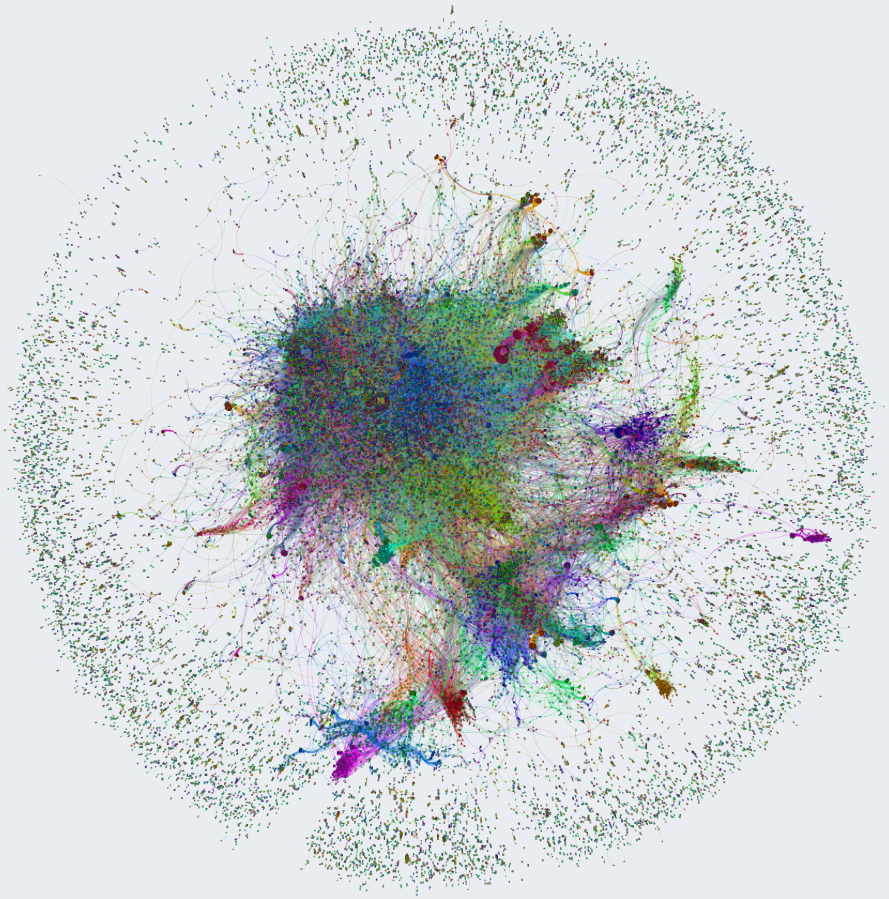
Resultados


Comparação dos métodos de particionamento
Comunidades detectadas corretamente (%)
Percentual relativo ao tamanho da rede inicial (%)
60
50
40
30
20
10
20
30
40
50
60
70
CNM
Label Propagation
Louvain
Leading Eigenvector
Walktrap

Resultados



Resultados



vértices: 69.837

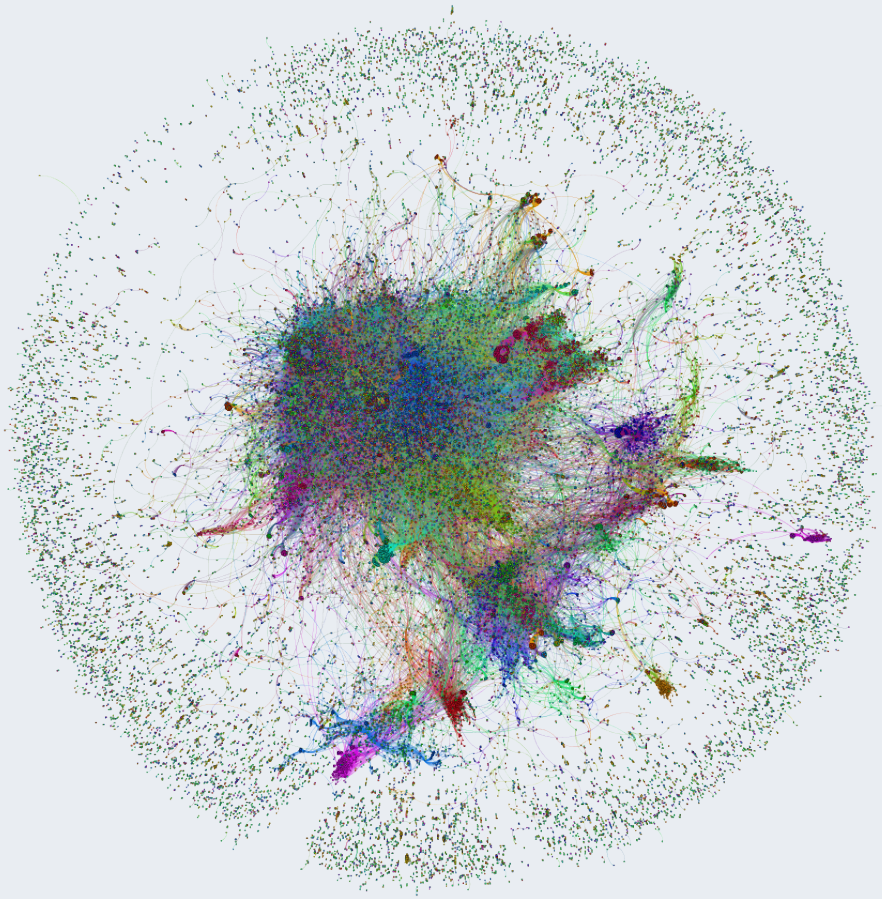
arestas: 134.700

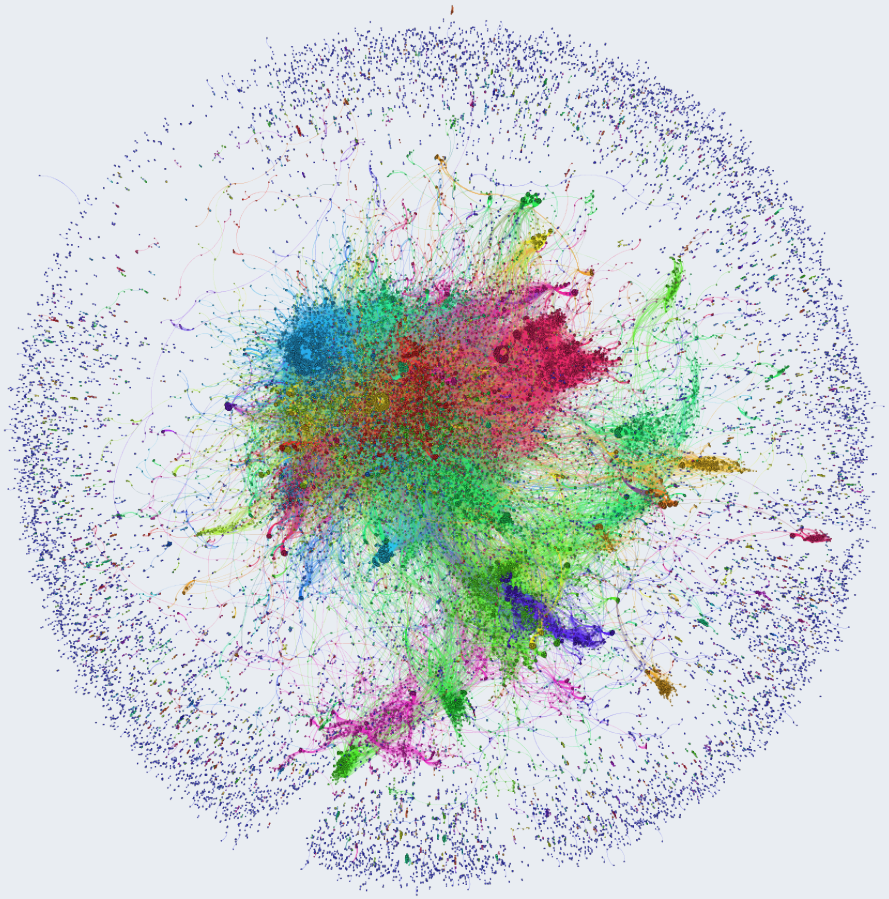
componentes conexas: 8.156

Tamanho da maior componente:

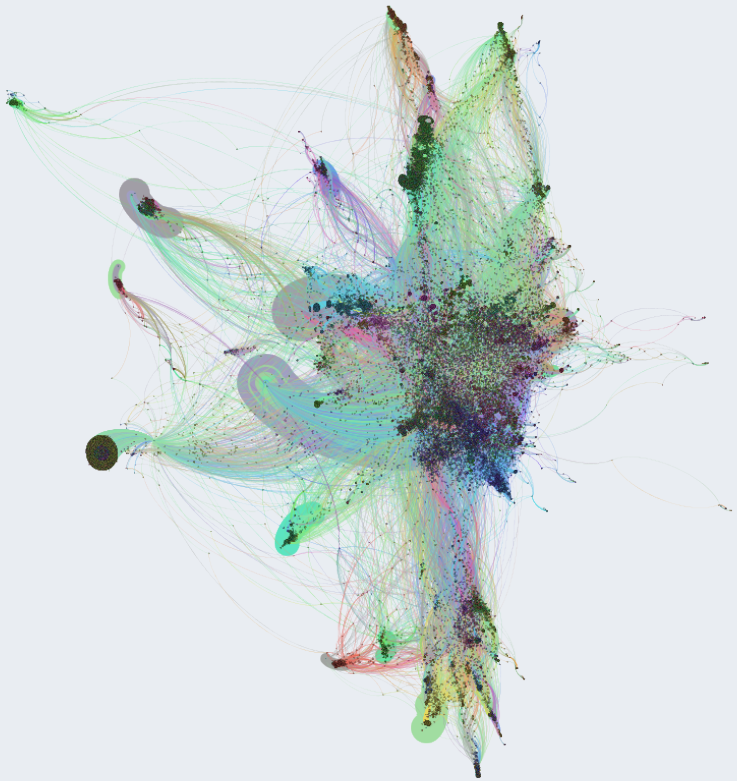
48.937

Resultados

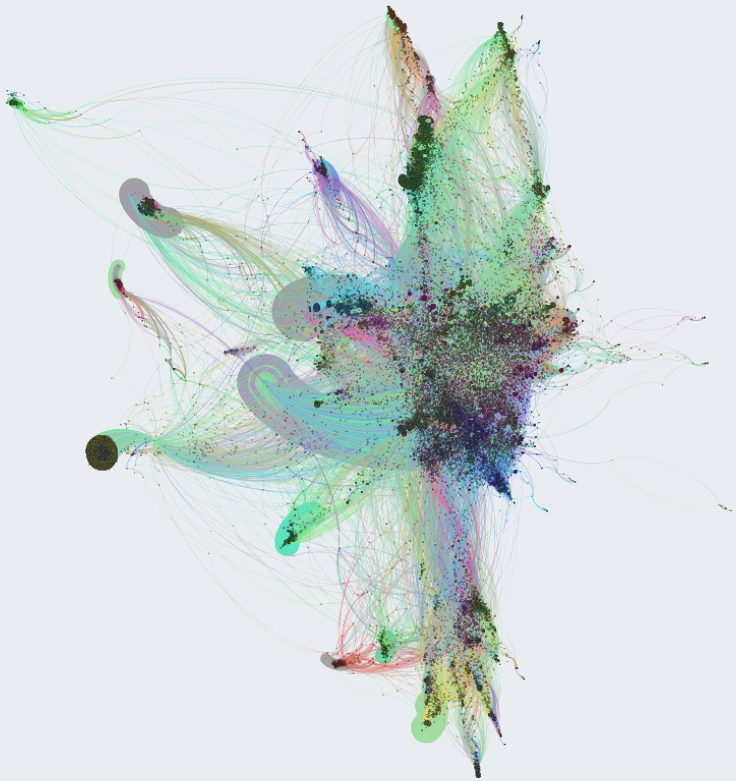




Resultados



Resultados



vértices: 70.132

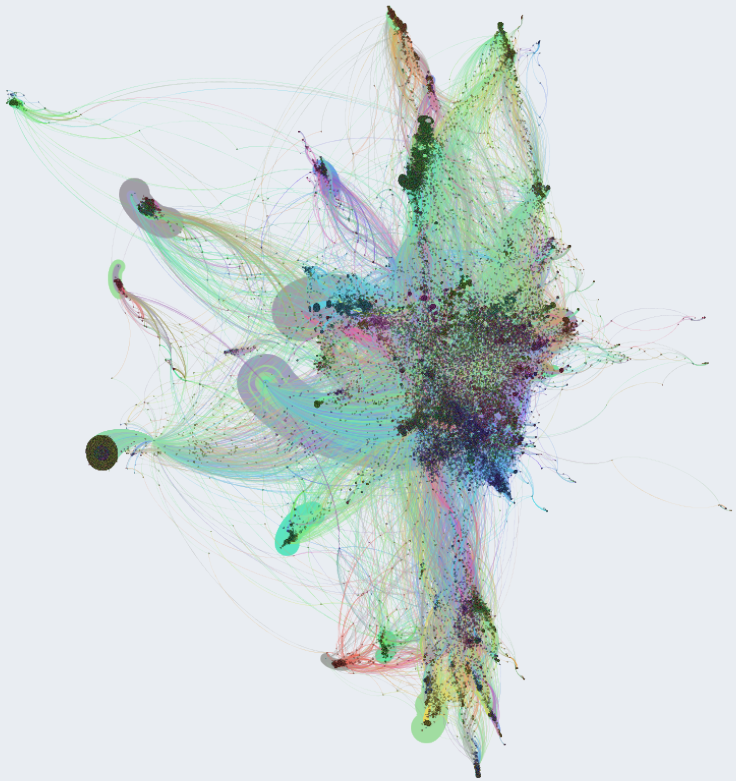
arestas: 404.100

componentes conexas: 6

Tamanho da maior componente:

69.765

Resultados



Considerações finais

Vantagens

- Alguns ótimos resultados na detecção

Considerações finais

Vantagens

- Alguns ótimos resultados na detecção

Desvantagens

- Tempo
- Estratégia de aplicar algoritmos mais custosos

Próximos passos

- Consolidar resultados

Próximos passos

- Consolidar resultados
- Análise qualitativa

Próximos passos

- Consolidar resultados
- Análise qualitativa
- Outras formas de detecção de comunidades (aprendizado de máquina)

Obrigado pela atenção

steventsukamoto@gmail.com