

Um Estudo sobre a Evolução Temporal de Comunidades Científicas

Bruno Leite Alves

Orientador: Alberto H. F. Laender

Coorientador: Fabrício Benevenuto

A sociedade está Organizada em Comunidades

- Existem vários tipos de comunidades
- Em uma rede social, indivíduos influenciam e são influenciados por outros indivíduos
- Comunidades têm líderes de opinião

Um grupo de líderes ou membros influentes são responsáveis por afetar o dinamismo de uma comunidade inteira

Trabalhos Relacionados

- Análises de estruturas de comunidades
[Ducheneaut et al., 2007; Kumar et al., 2006; Patil et al., 2012]
- Investigação sobre as evoluções de redes
[Leskovec et al., 2005, 2008; Viswanath et al., 2009]
- Caracterização de comunidades científicas
[Backstrom et al., 2006; Huang et al., 2008]
- Modelos de geração de grafos
[Ducheneaut et al., 2007; Kumar et al., 2006; Leskovec et al., 2005, 2008]
- Extração do núcleo com base nas propriedades estruturais da rede
[Chakrabarti et al., 2006; Hopcroft et al., 2004; Leskovec et al., 2010, Sachan et al., 2012; Seifi et al., 2012]

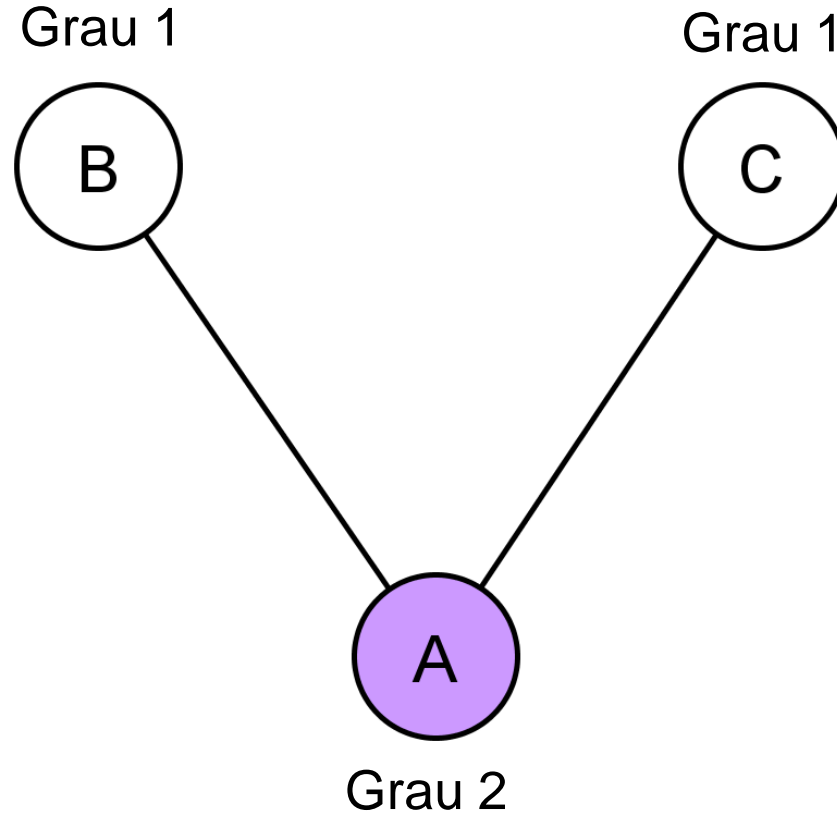
Contribuições

- Definição de uma métrica capaz de quantificar a importância de um pesquisador
- Descrição de uma estratégia capaz de estimar o índice h de um pesquisador ao longo do tempo
- Definição do conceito de núcleo de uma comunidade a partir da métrica proposta
- Caracterização de comunidades científicas e discussões de como nossa métrica afeta as propriedades das redes
- Visualização das comunidades estudadas

Redes Complexas

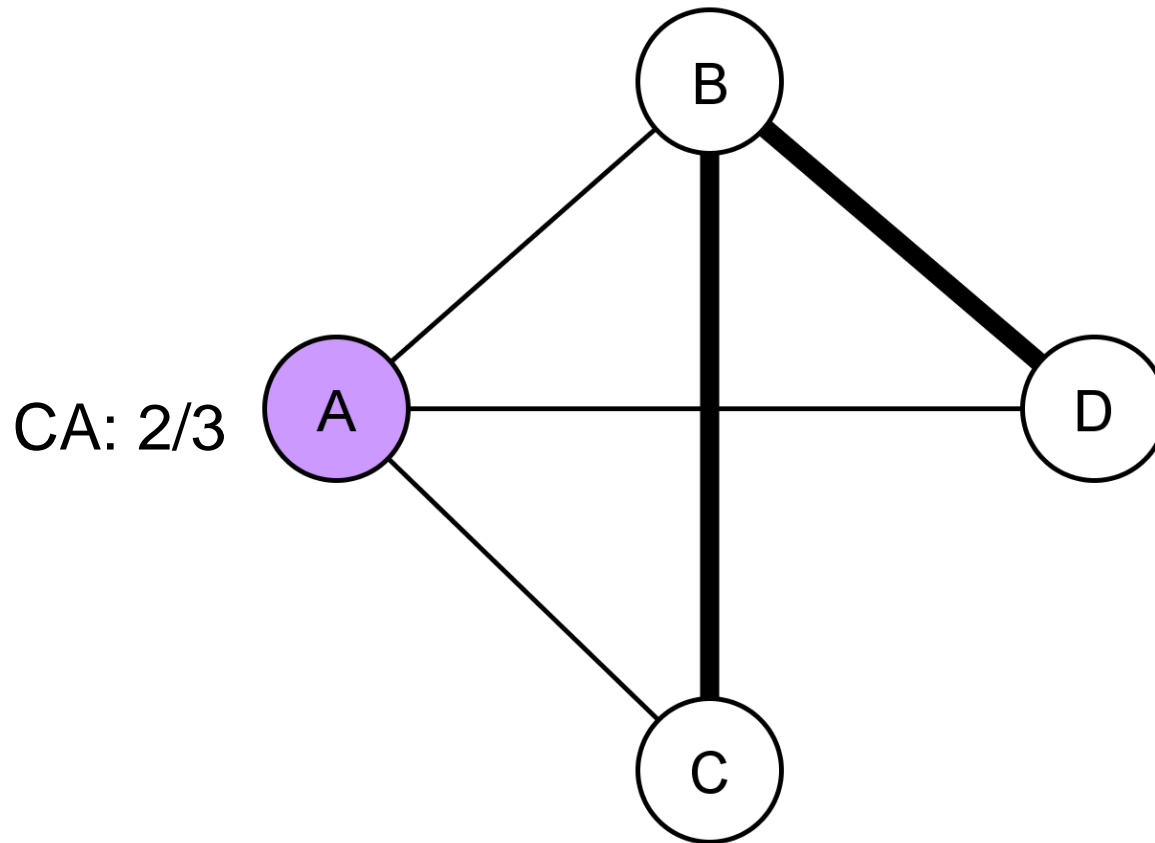
- Uma rede é um conjunto de itens conectados entre si
- Uma rede pode ser modelada como um grafo que possui um conjunto de nodos e arestas
- *Redes sociais online* possuem características que nos permitem modelá-las como redes
- Redes de colaboração científica são formadas por pesquisadores que publicam trabalhos em fóruns científicos

Grau dos Nodos



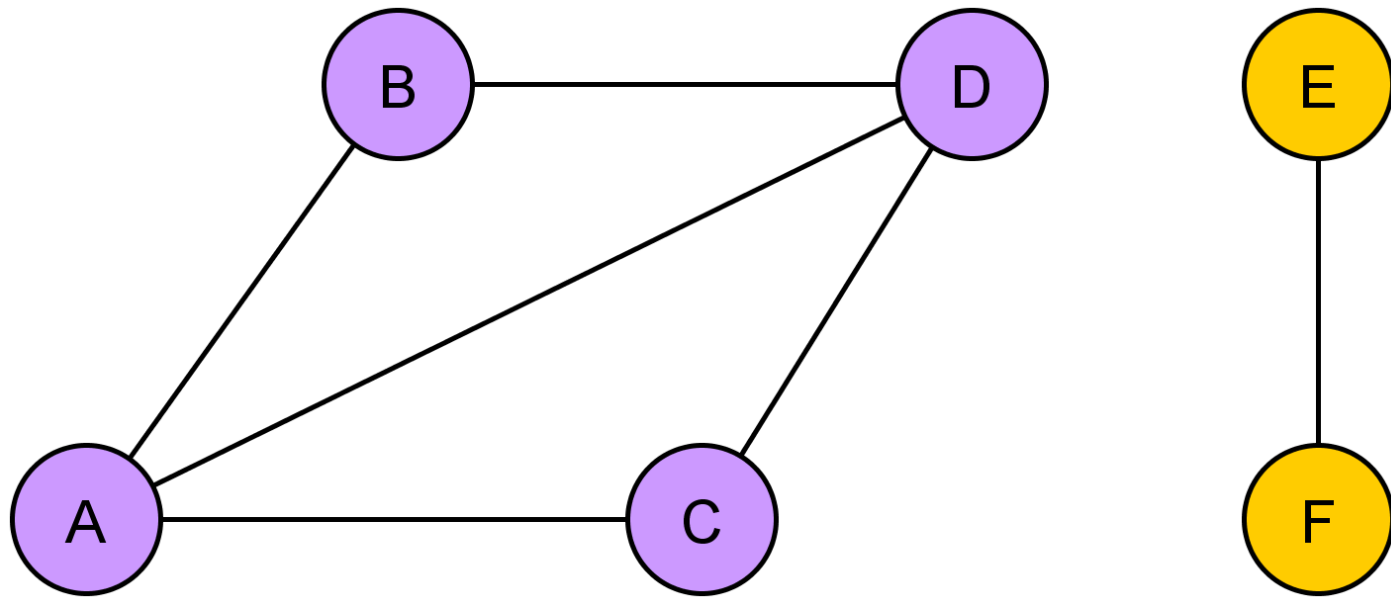
Número de arestas incidentes àquele nodo

Coeficiente de Agrupamento



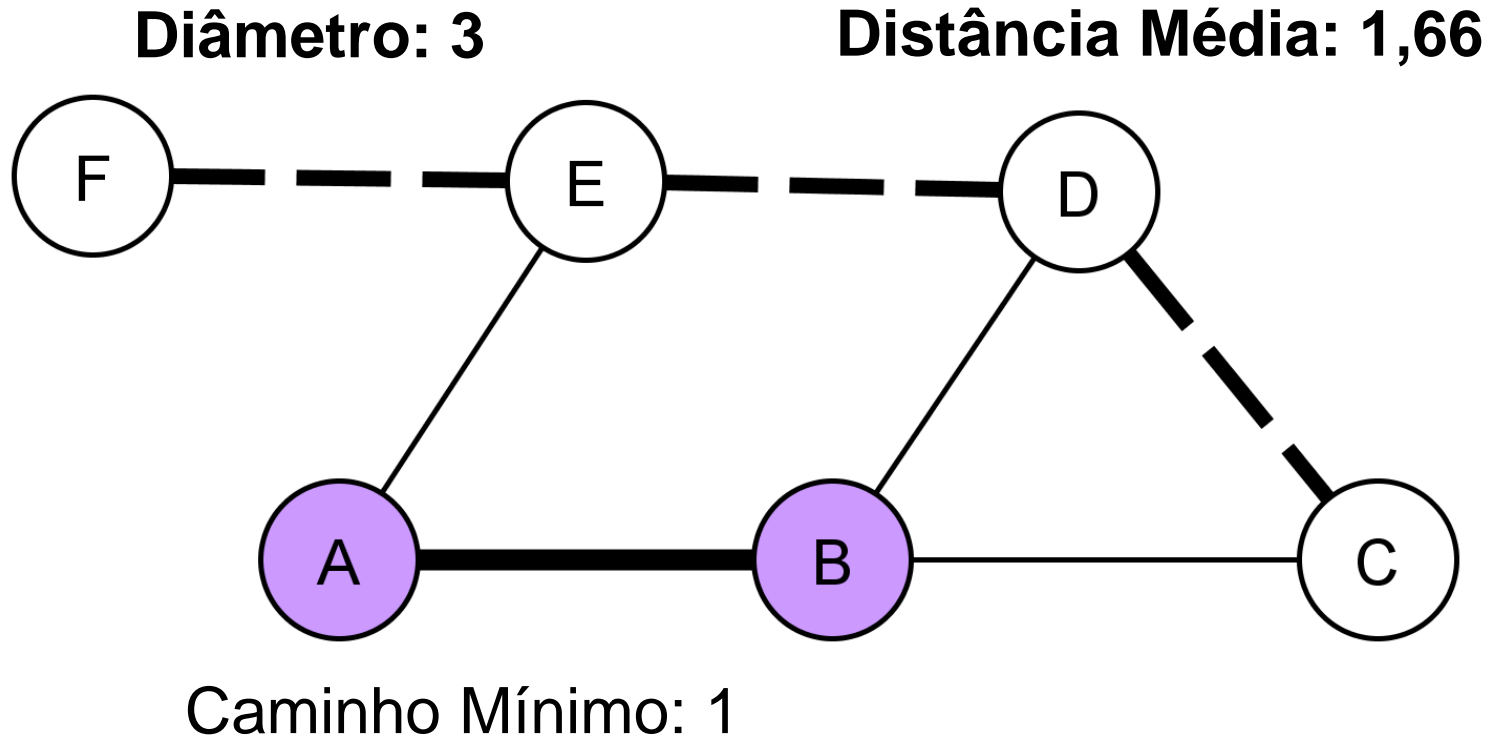
O CA informa o quão agrupados os vizinhos de um dado nodo se encontram na rede

Componentes



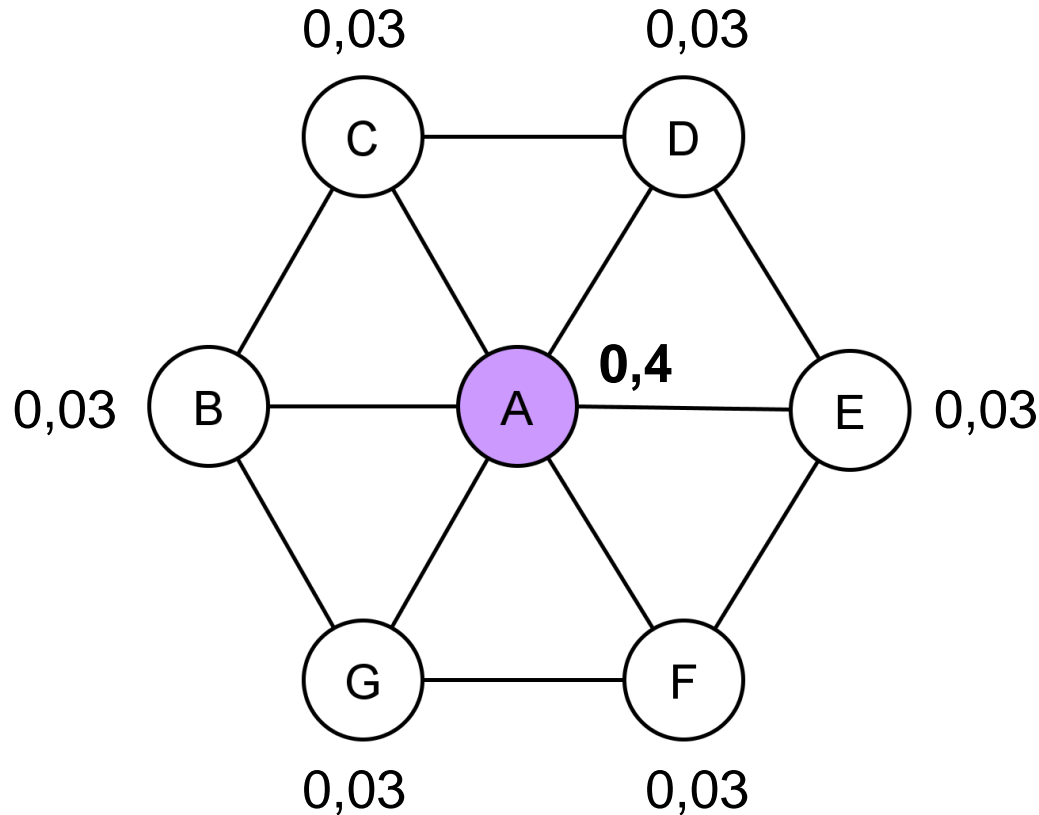
Um componente é um subconjunto de nodos interligados entre si

Distância Média e Diâmetro



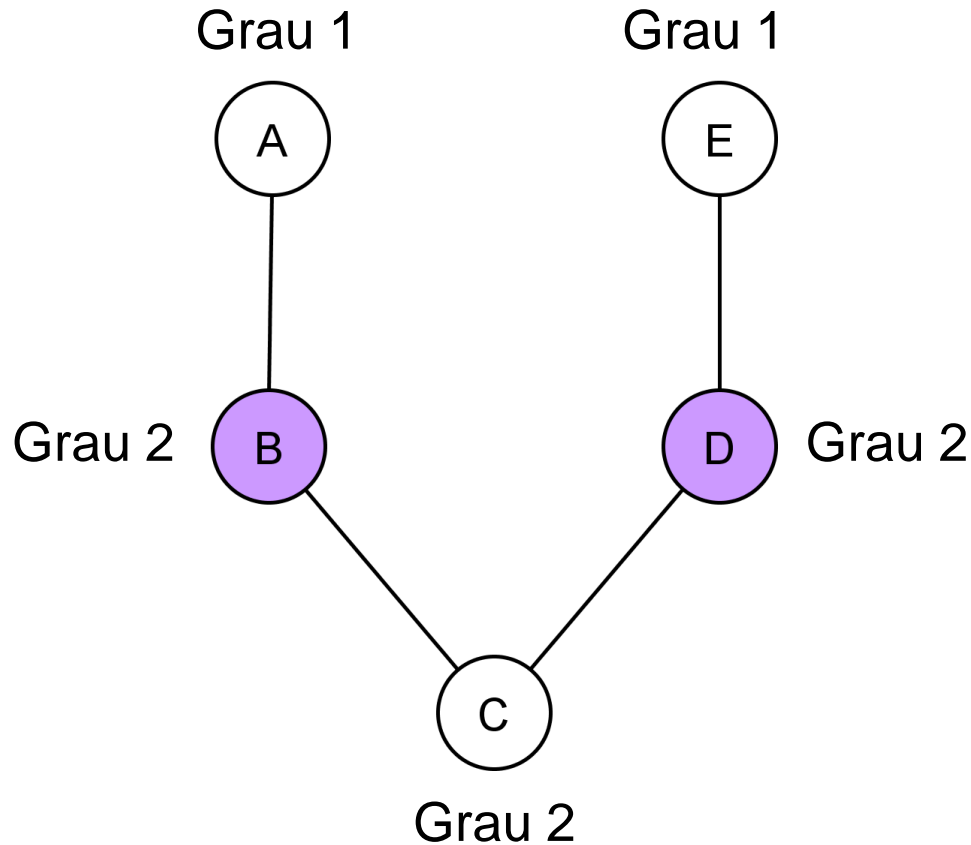
A distância média é a média do número de arestas em todos os caminhos mínimos existentes

Betweenness



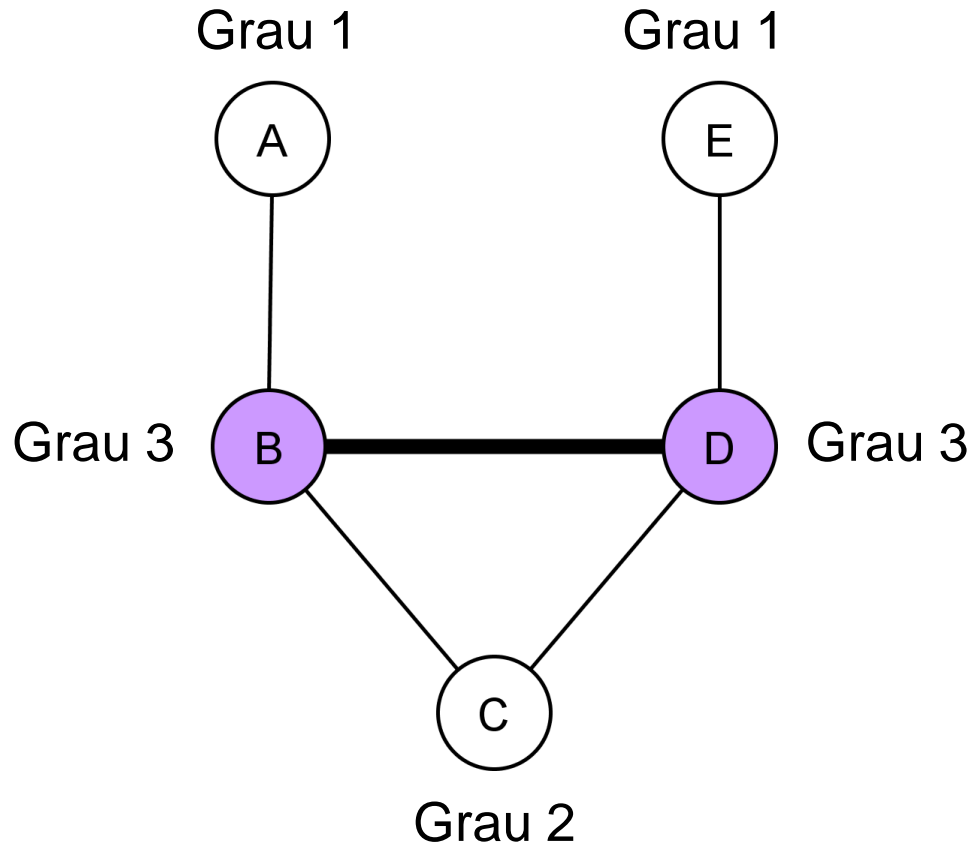
Mede a importância de um nodo considerando o número de caminhos mínimos que por ele passam

Assortatividade



Indica se os nodos tendem a estabelecer conexões com outros nodos de mesmo grau ou não

Assortatividade



Indica se os nodos tendem a estabelecer conexões com outros nodos de mesmo grau ou não

Comunidades Científicas



2,2 milhões de publicações
de **1,2 milhões de autores**

22 conferências
dos **maiores SIGs da ACM**

Consideramos cada **conferência**
como uma **comunidade científica**

SIGs da ACM Considerados

SIGACT	SIGDOC	SIGMOD
SIGAPP	SIGGRAPH	SIGOPS
SIGARCH	SIGIR	SIGPLAN
SIGBED	SIGKDD	SIGSAC
SIGCHI	SIGMETRICS	SIGSAM
SIGCOMM	SIGMICRO	SIGSOFT
SIGCSE	SIGMM	SIGUCCS
SIGDA	SIGMOBILE	SIGWEB

CoScore

- Estima a importância de um pesquisador dentro da comunidade
- O *CoScore* de um pesquisador p com um índice h em uma comunidade c em um período de tempo t é dado por:

$$CoreScore_{p,c,t} = h_{p,t} \times \#publicações_{p,c,t}$$

Como estimar o Índice H

Somente **30%**
dos autores da DBLP possuem um perfil no
Google Scholar

Como estimar o Índice H

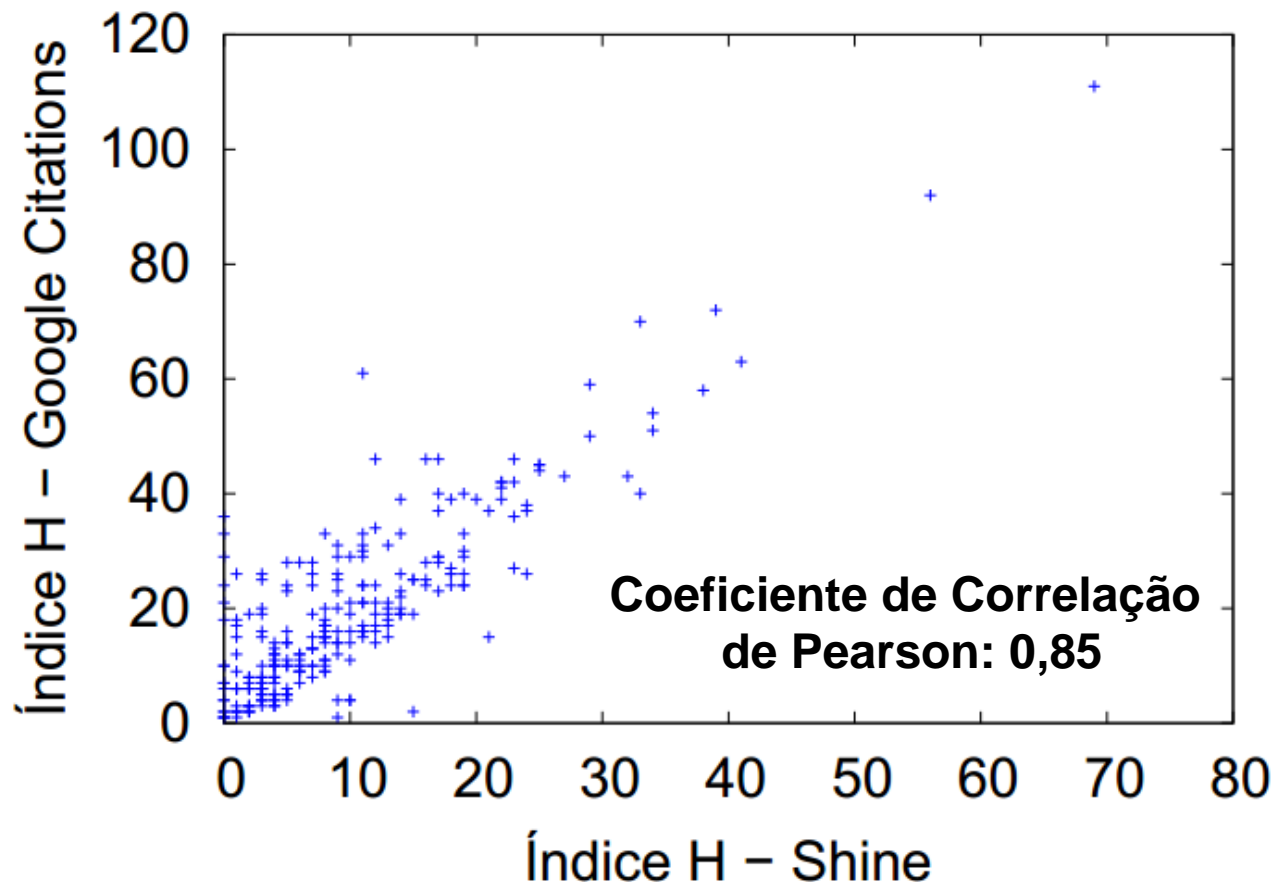
Somente **30%**
dos autores da DBLP possuem um perfil no
Google Scholar

Alternativa:



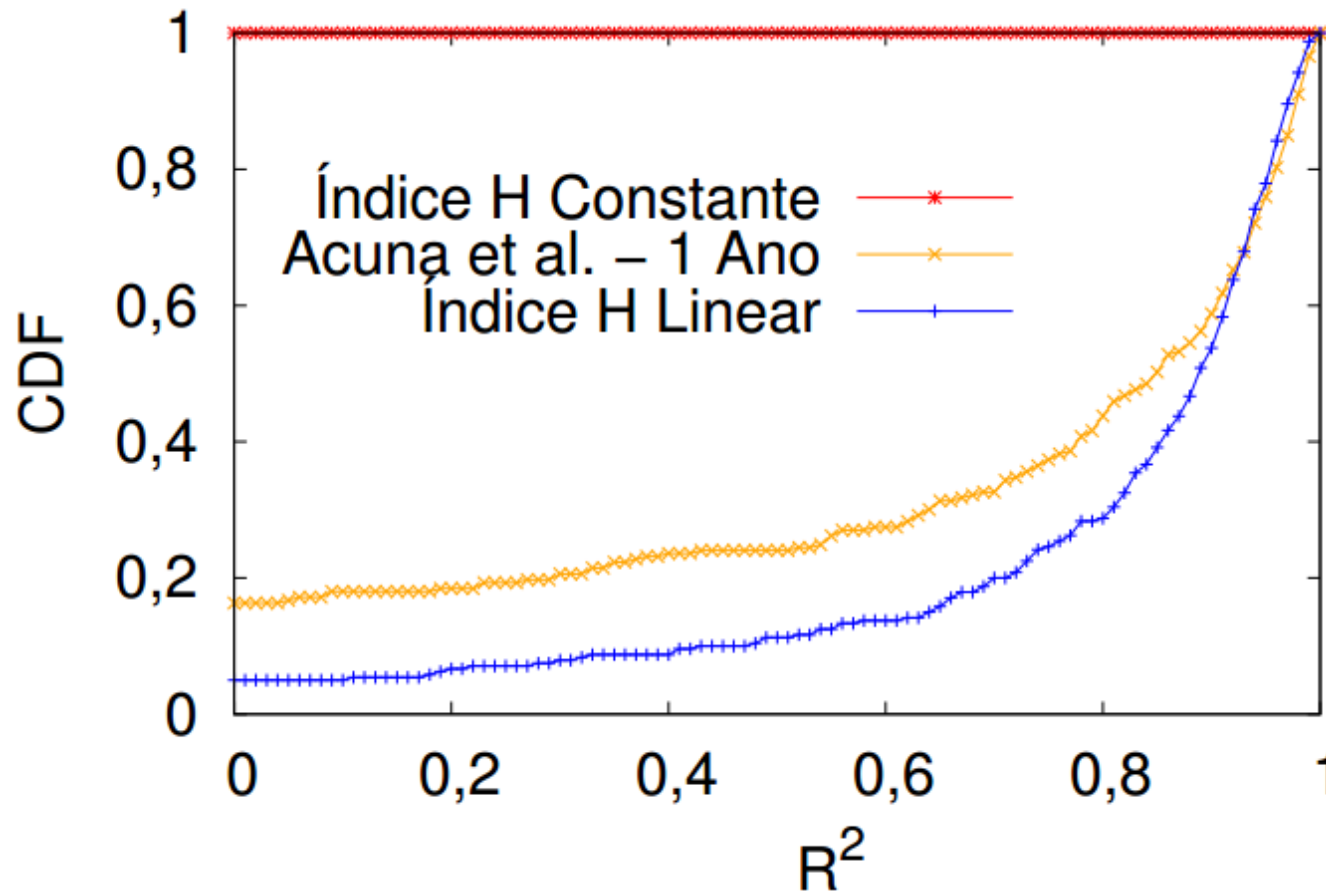
www.shine.icomp.ufam.edu.br

Shine vs. Google Scholar



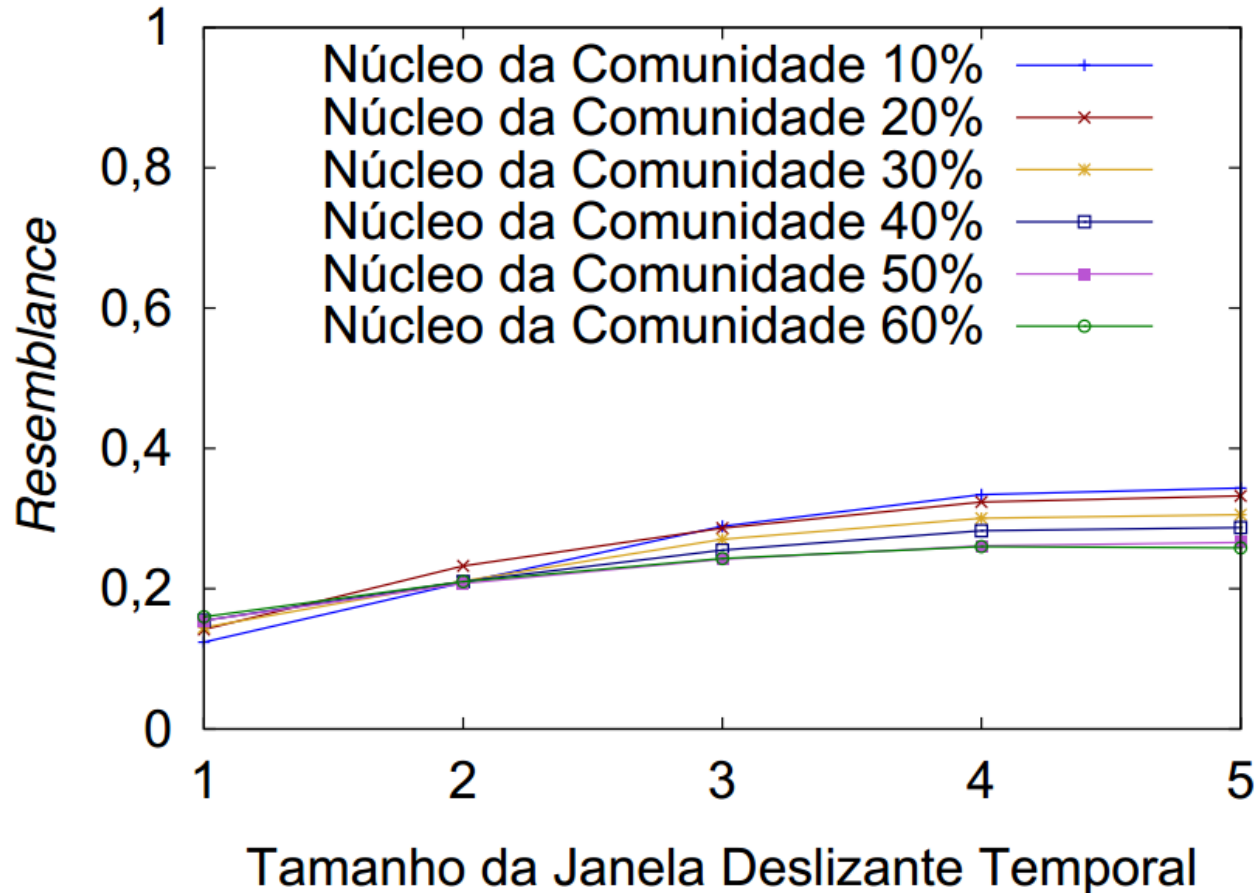
Shine oferece uma boa estimativa para o índice h

Evolução do Índice H



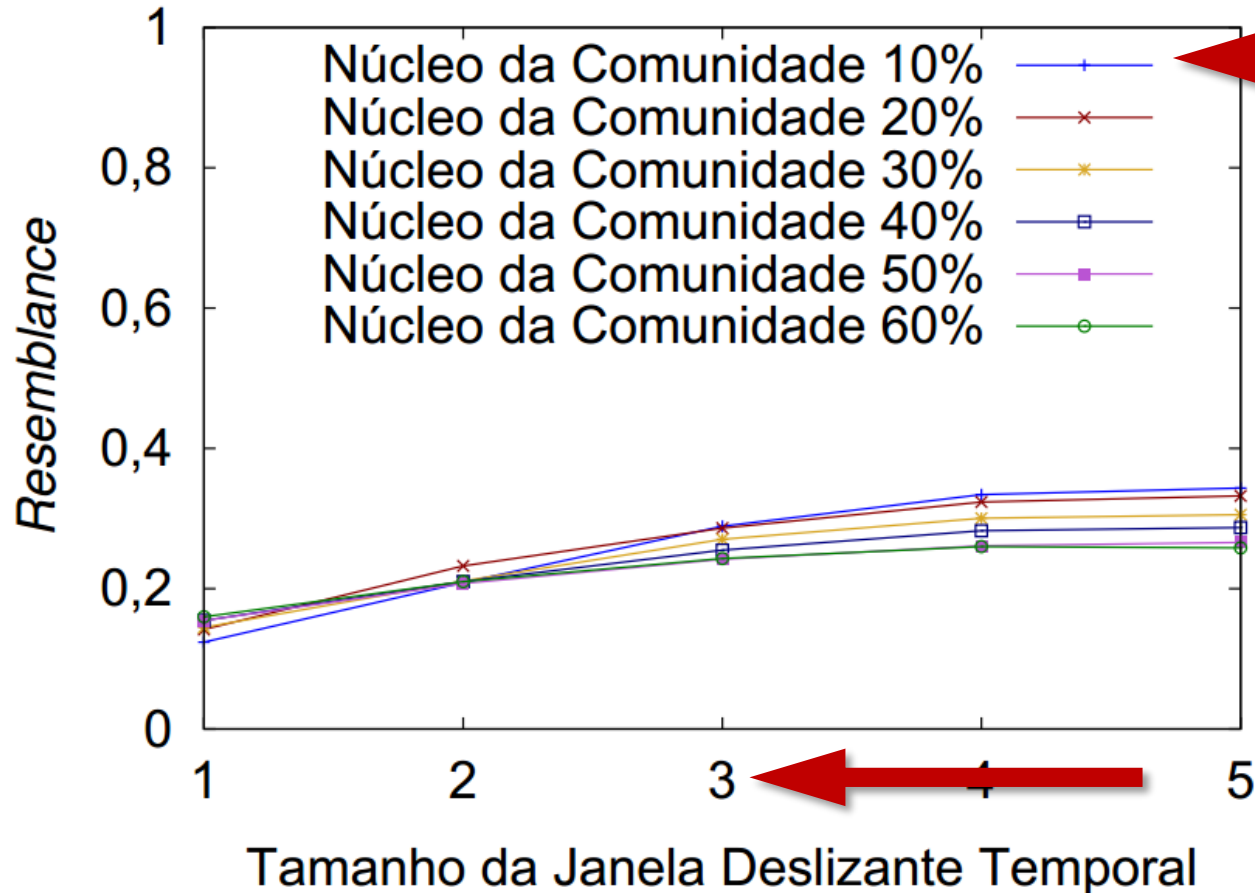
O índice h tende, no geral, a manter uma evolução linear

Resemblance e Coeficiente Angular



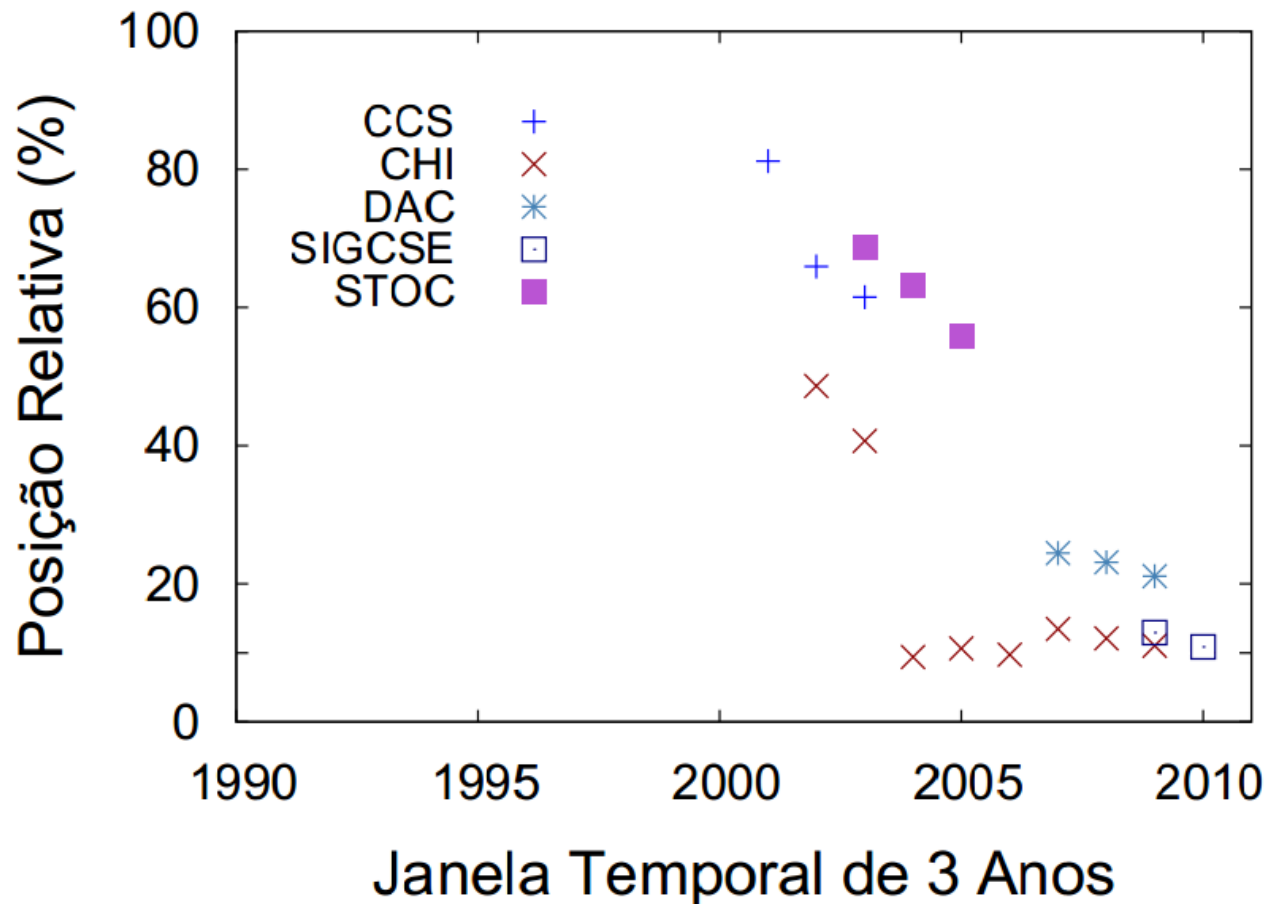
SIGMOD

Resemblance e Coeficiente Angular



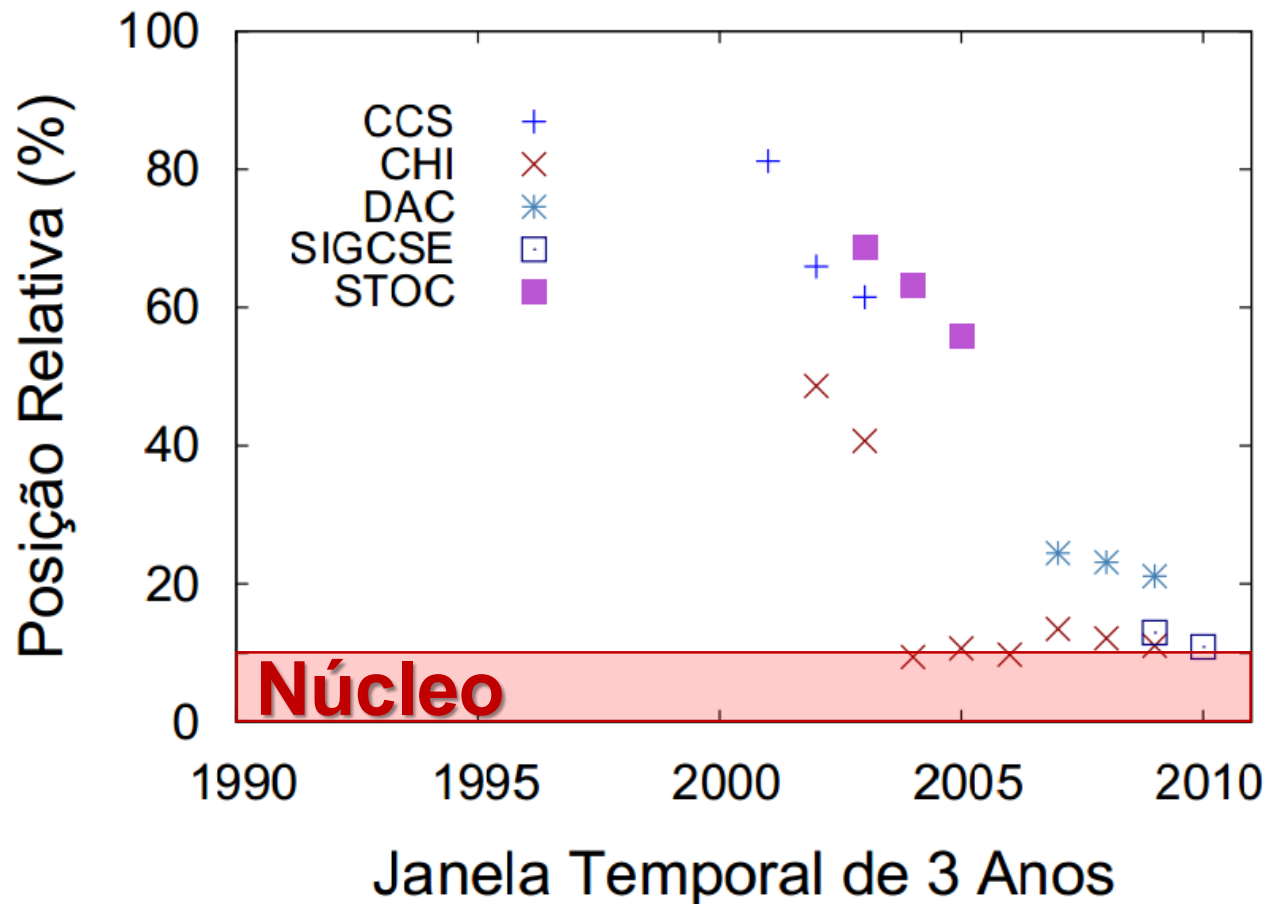
SIGMOD

Comunidades de Luis von Ahn



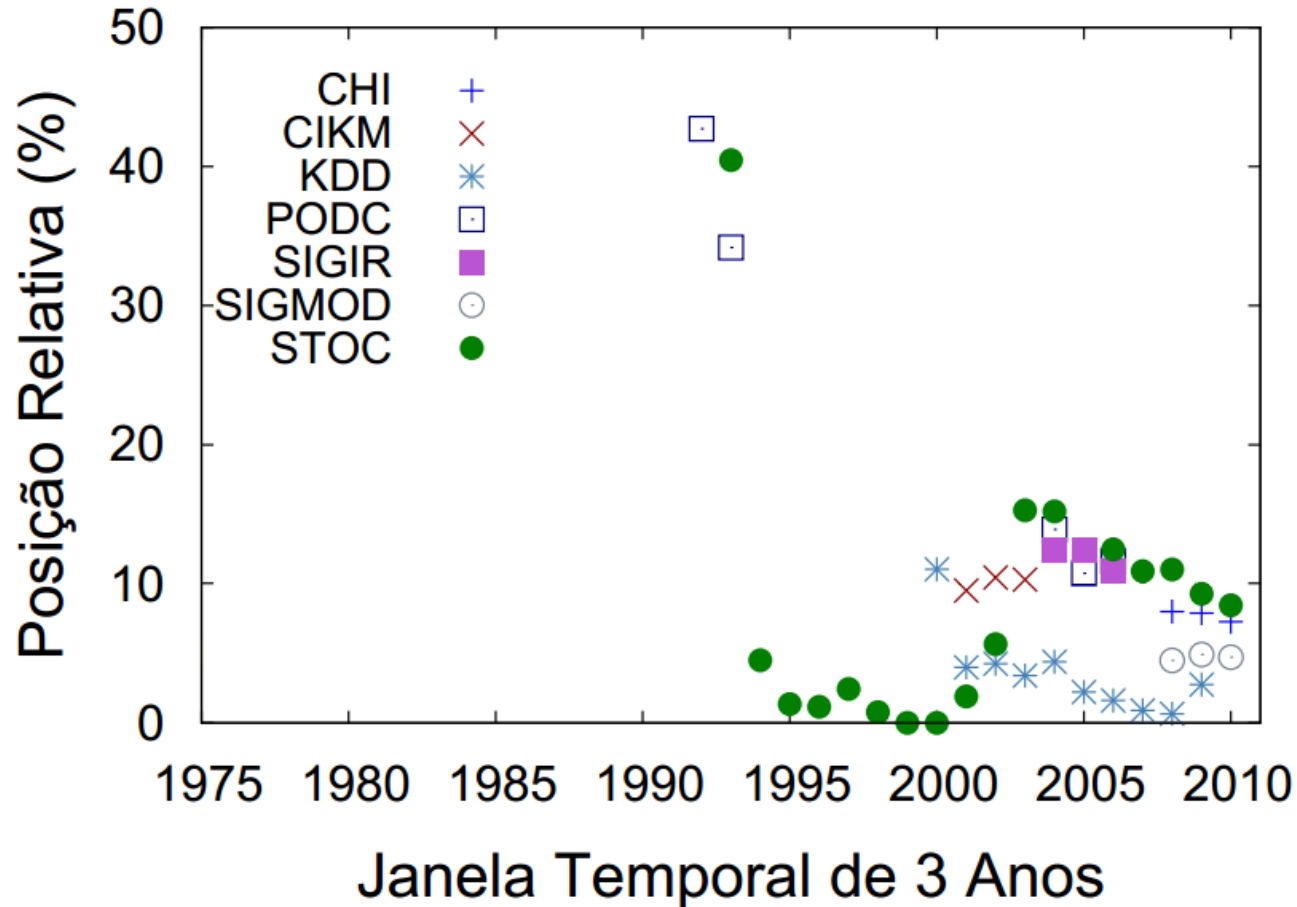
Alto CoScore nas comunidades CHI e SIGCSE

Comunidades de Luis von Ahn



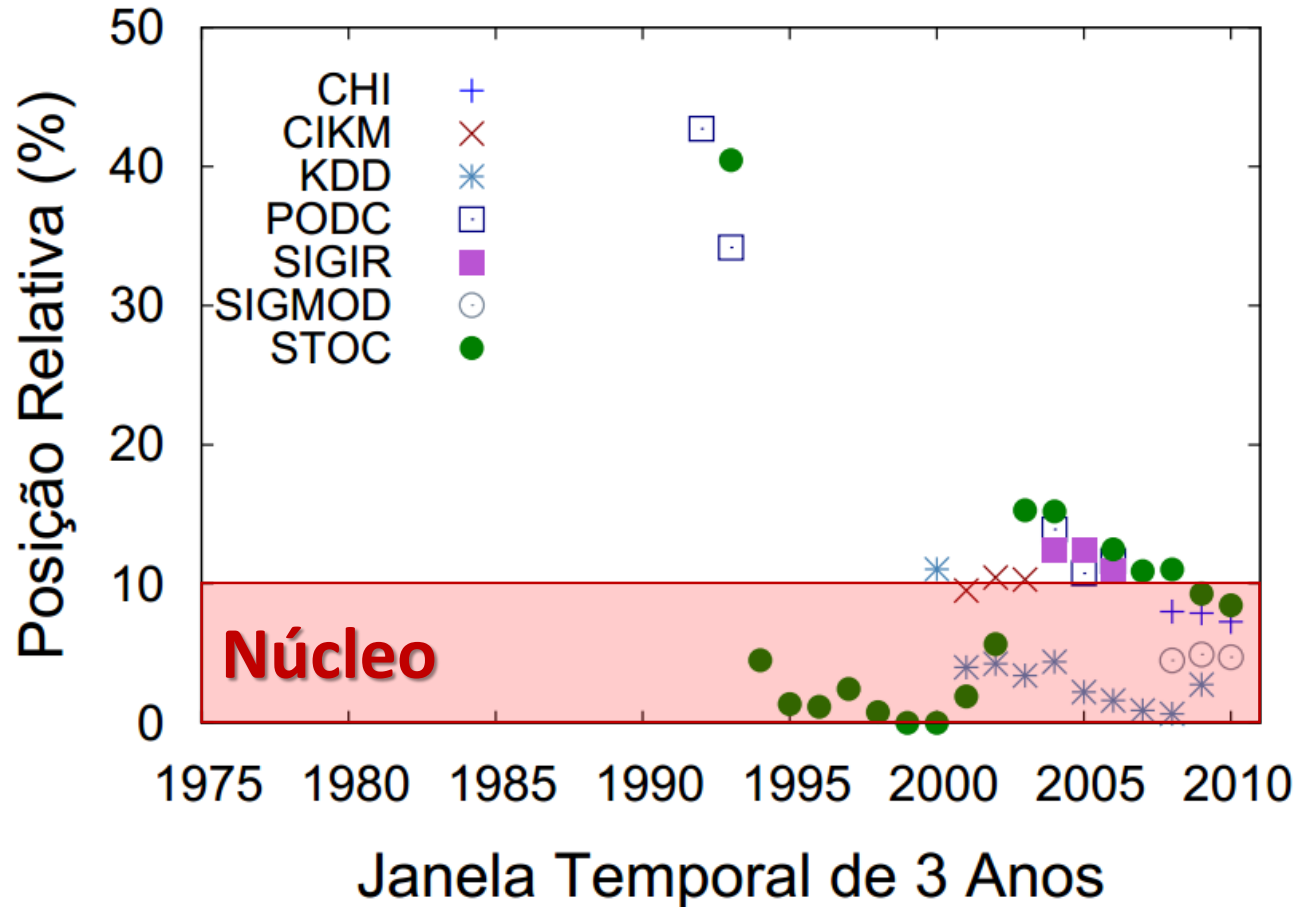
Membro do núcleo da CHI

Comunidades de Jon Kleinberg



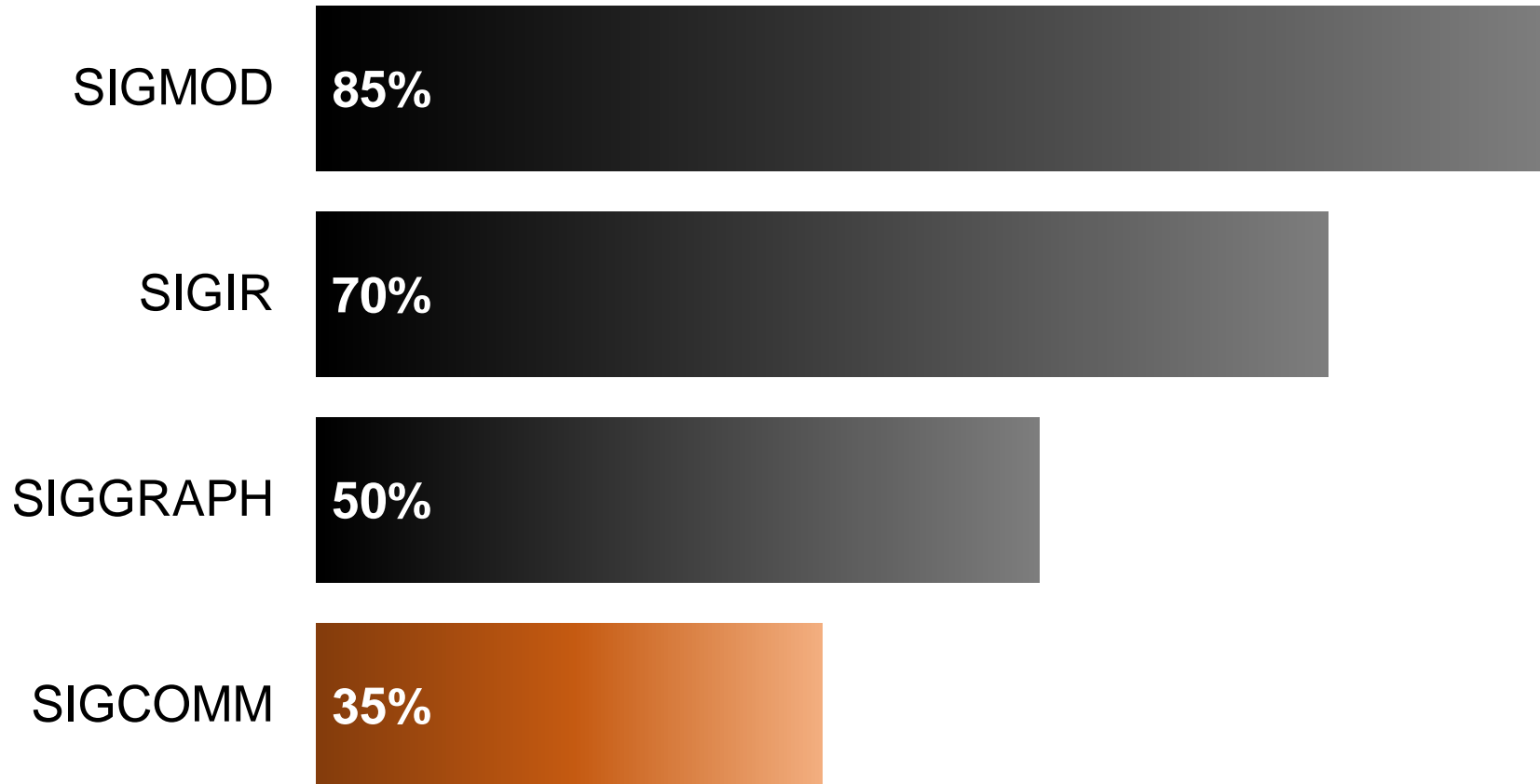
Alto CoScore em várias comunidades

Comunidades de Jon Kleinberg



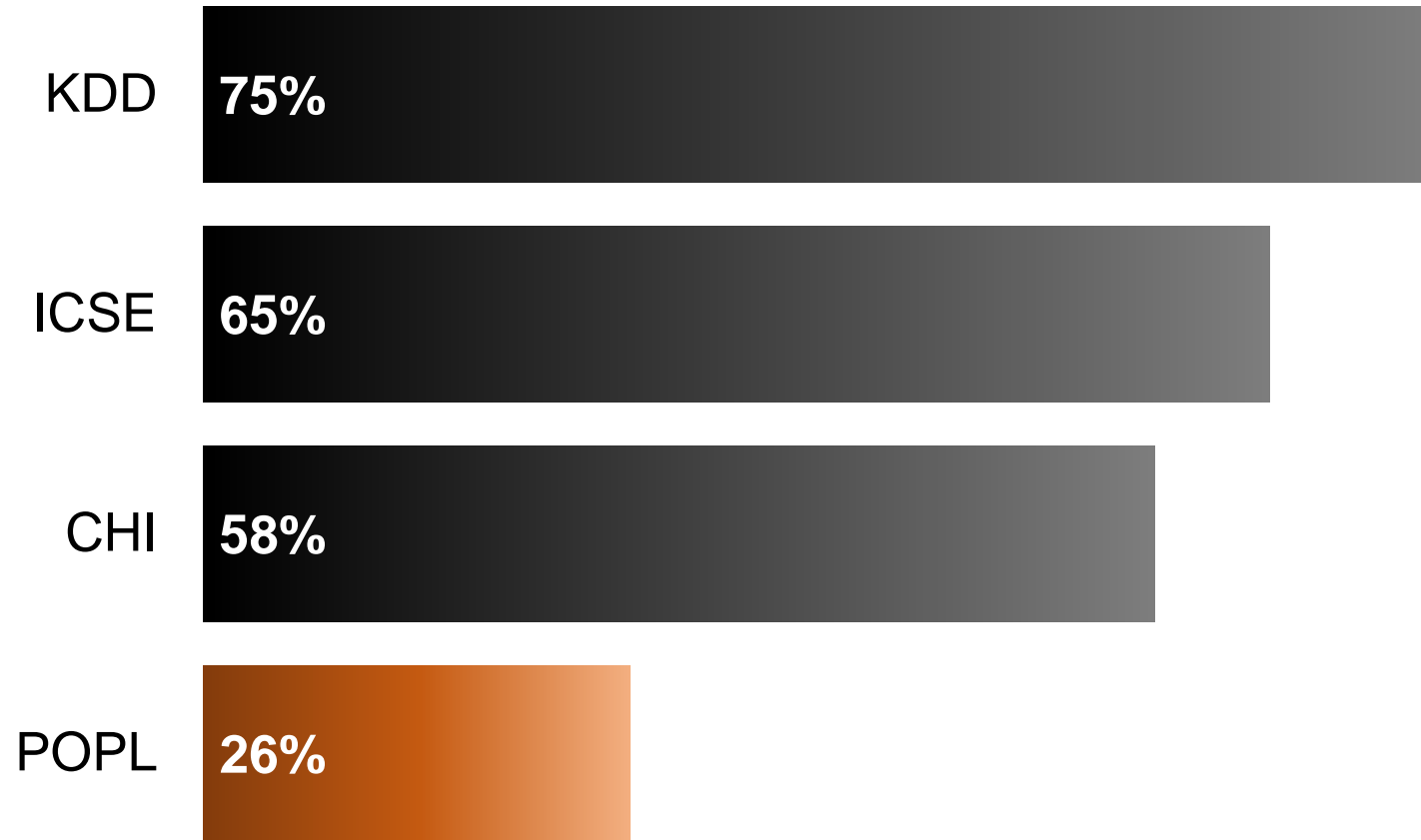
Mudança da STOC para KDD

Pesquisadores Premiados



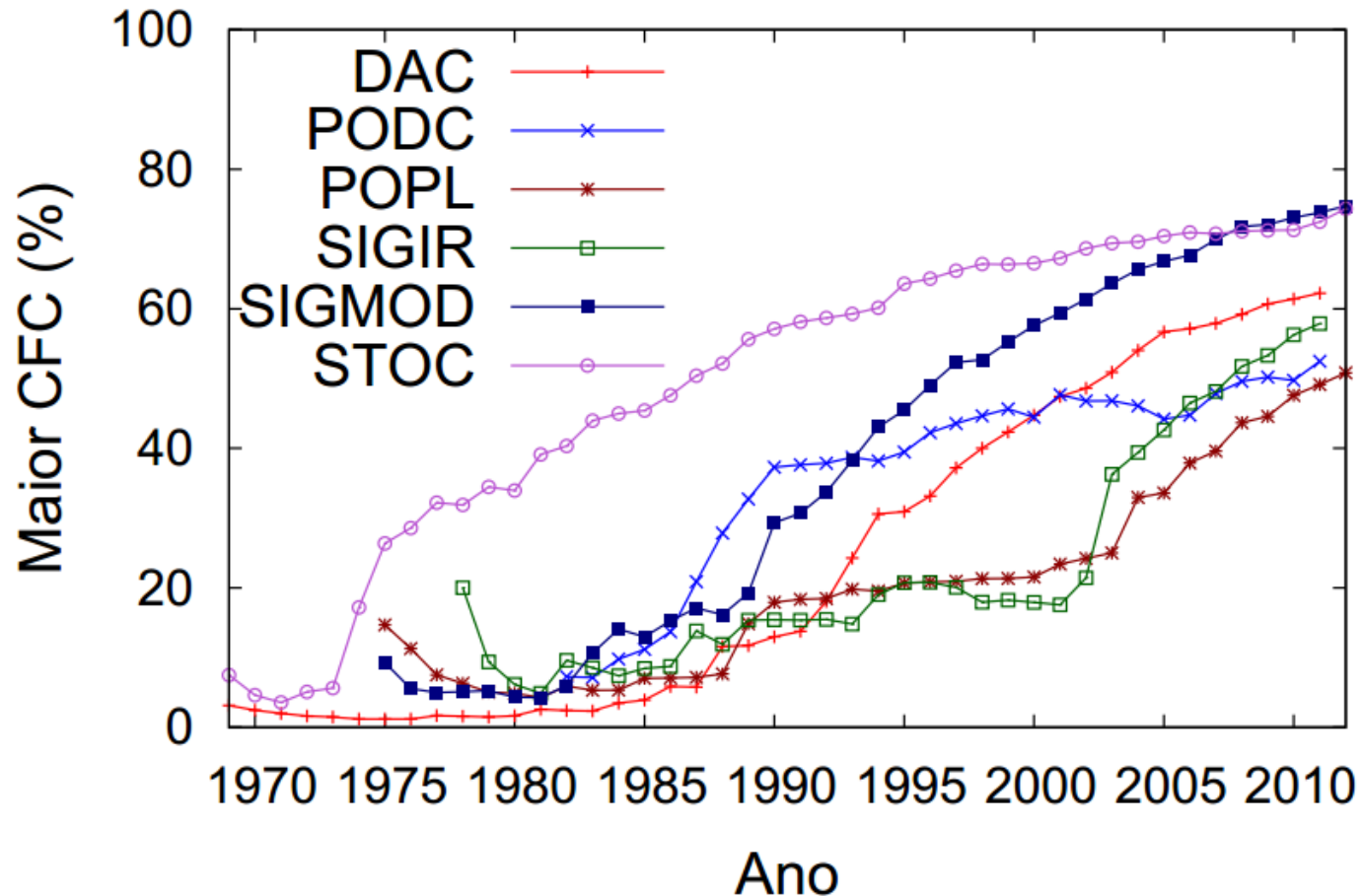
Membros dos núcleos das comunidades que receberam prêmios

Pesquisadores Premiados



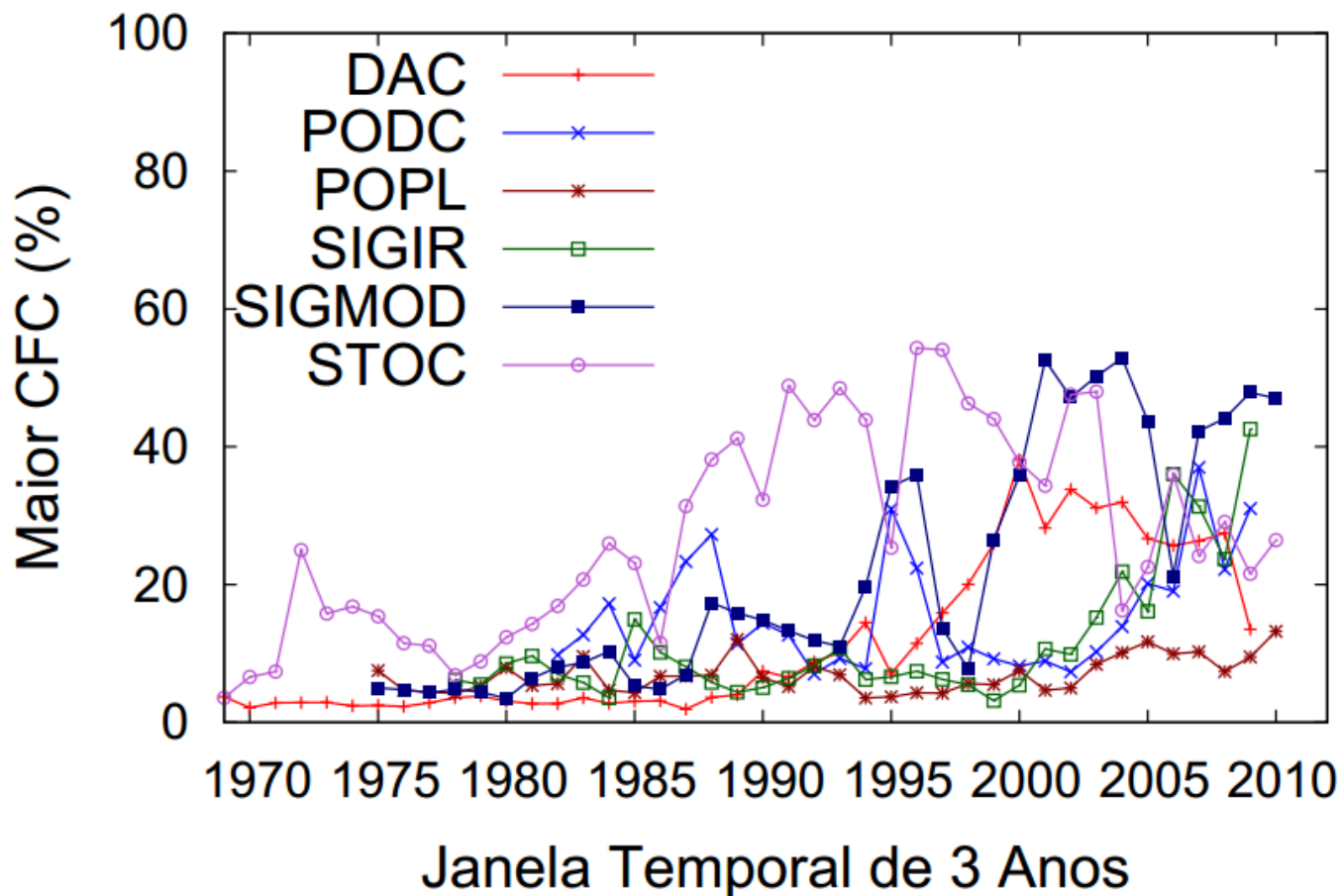
Membros da POPL foram premiados com o *ACM A.M. Turing Award*

Evolução das Comunidades Científicas



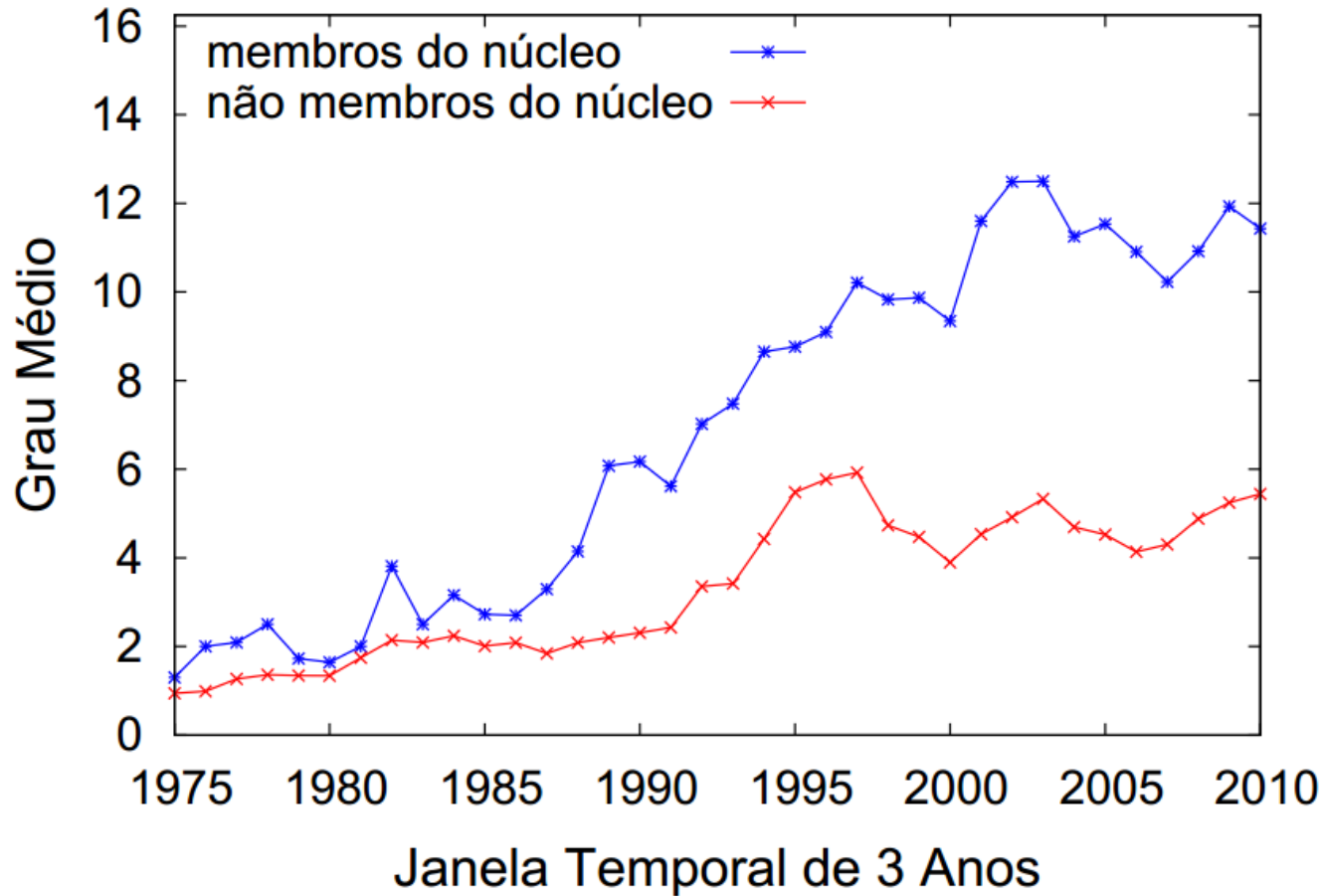
Ano a ano acumulando nodos e arestas

Evolução das Comunidades Científicas



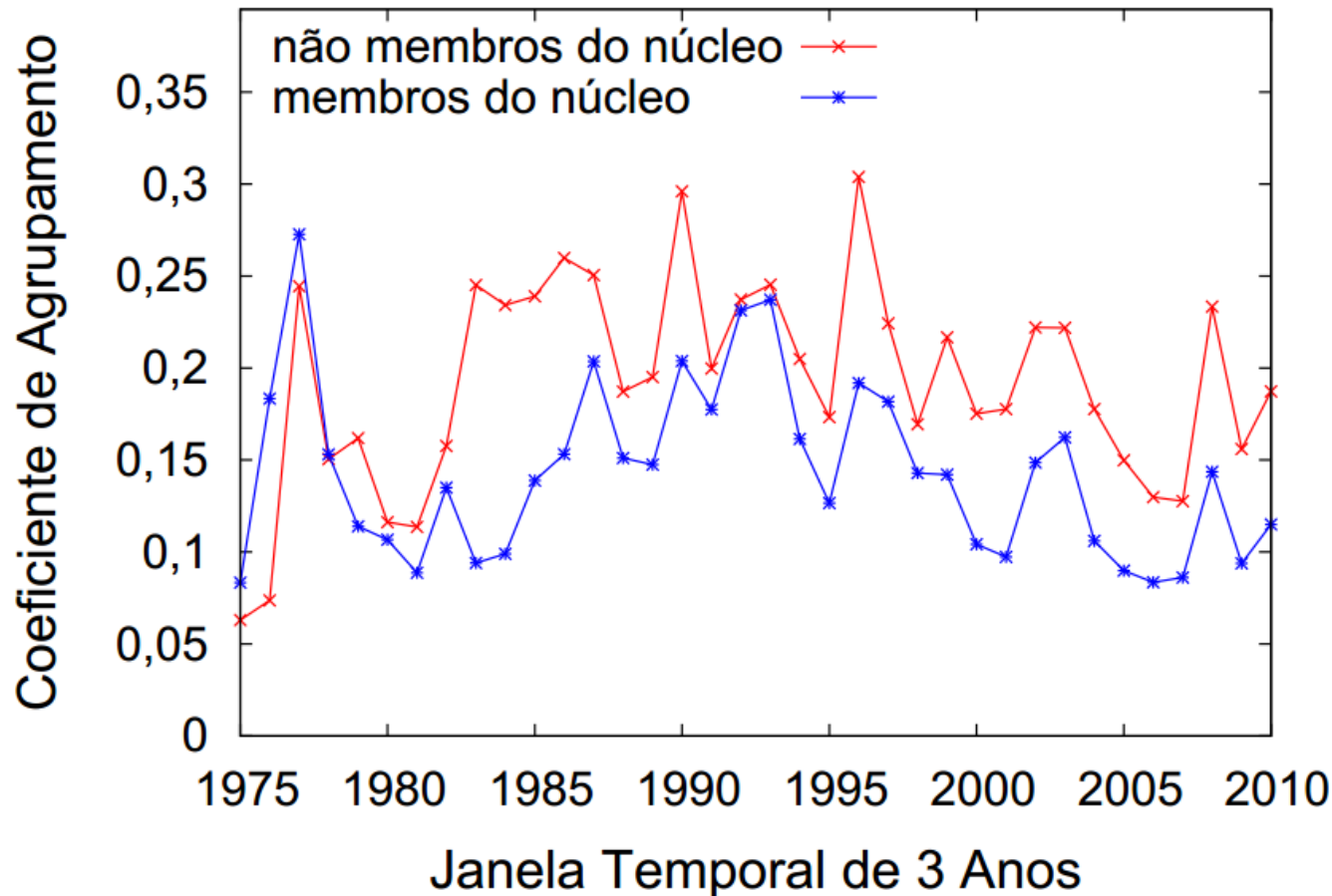
Instância (3 anos) construída com base em nodos e arestas

Membros do Núcleo vs. Não Membros



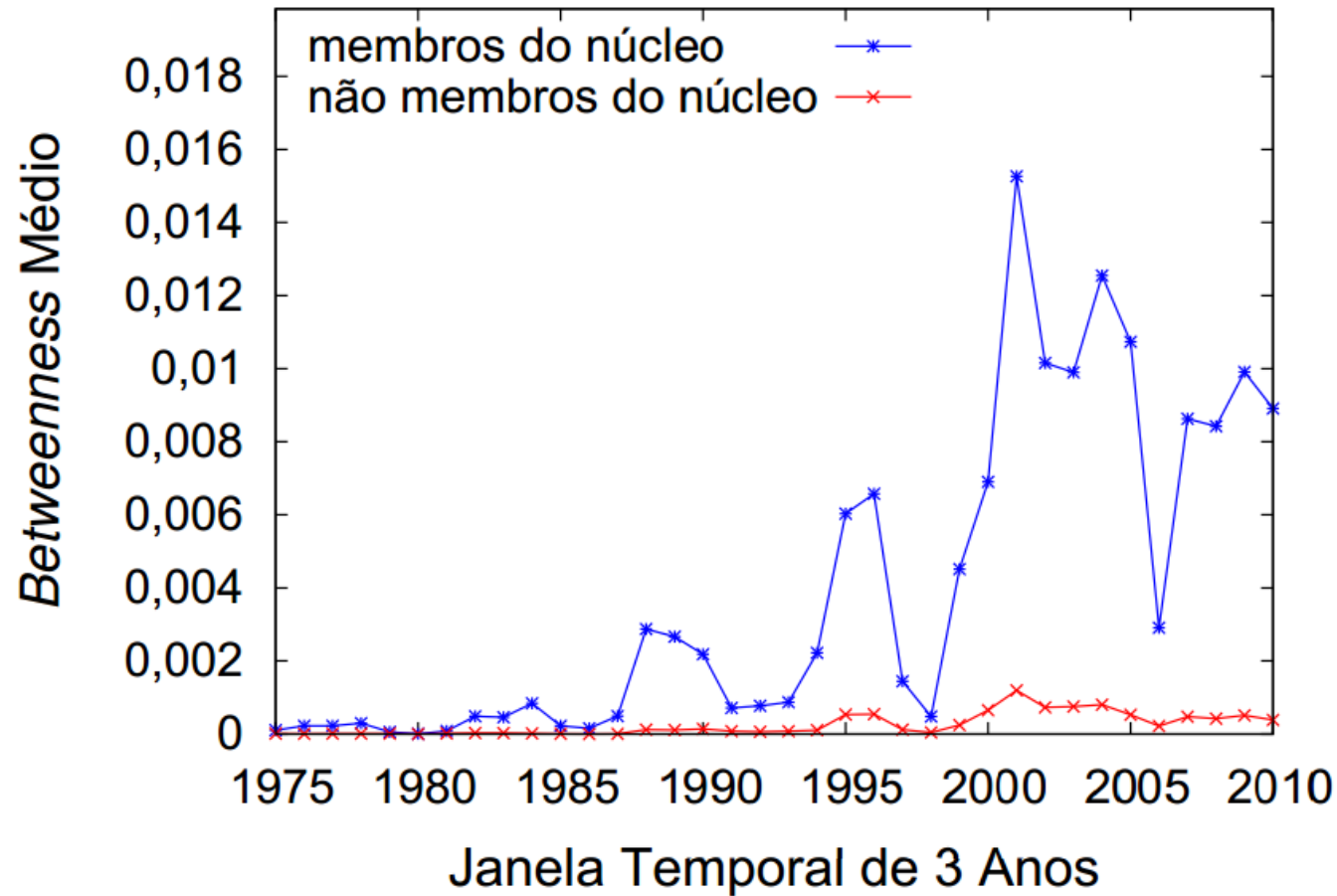
O grau médio dos membros do núcleo é maior que os dos não membros

Membros do Núcleo vs. Não Membros



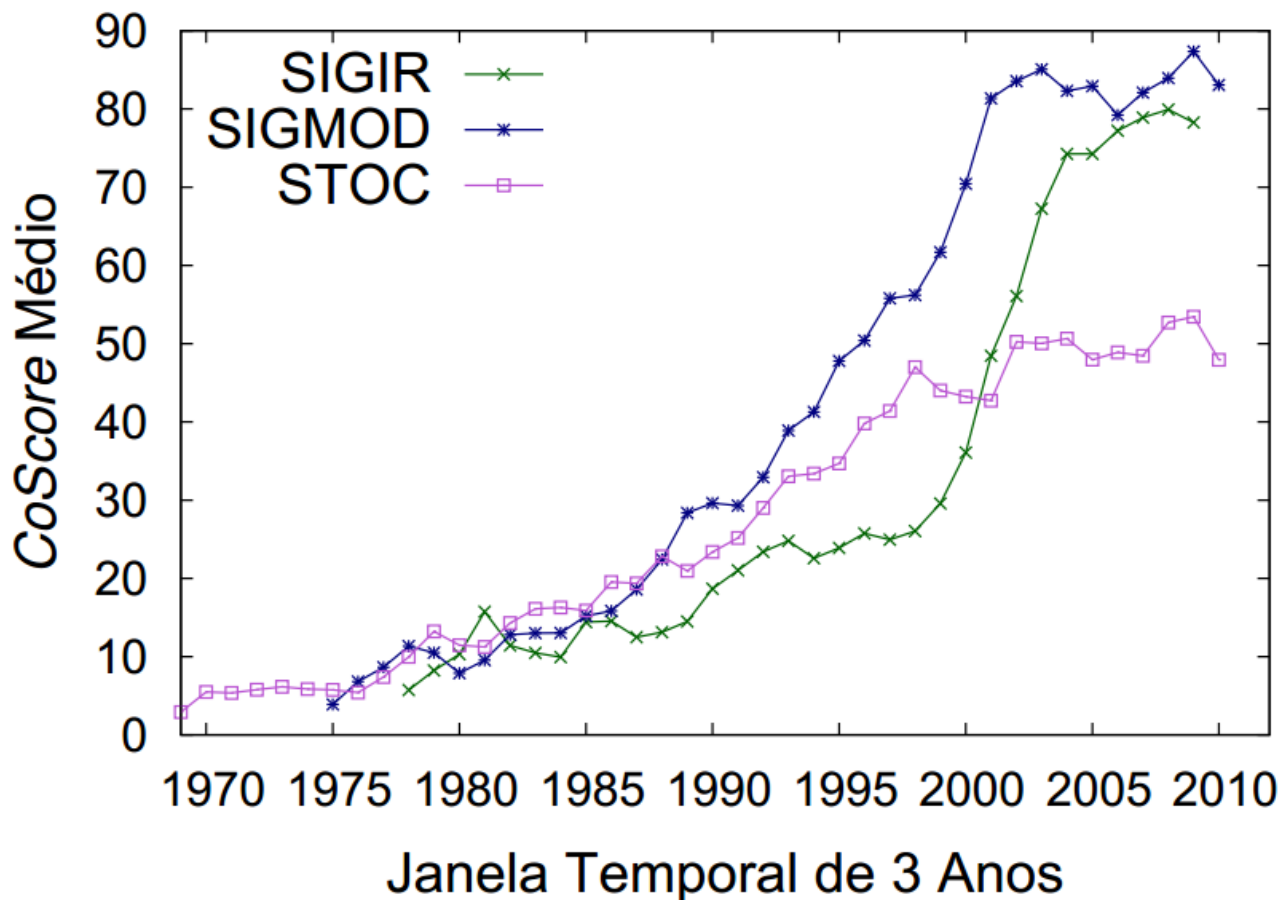
O núcleo pode atuar como *hub*, conectando diferentes grupos com pequenas interseções

Membros do Núcleo vs. Não Membros



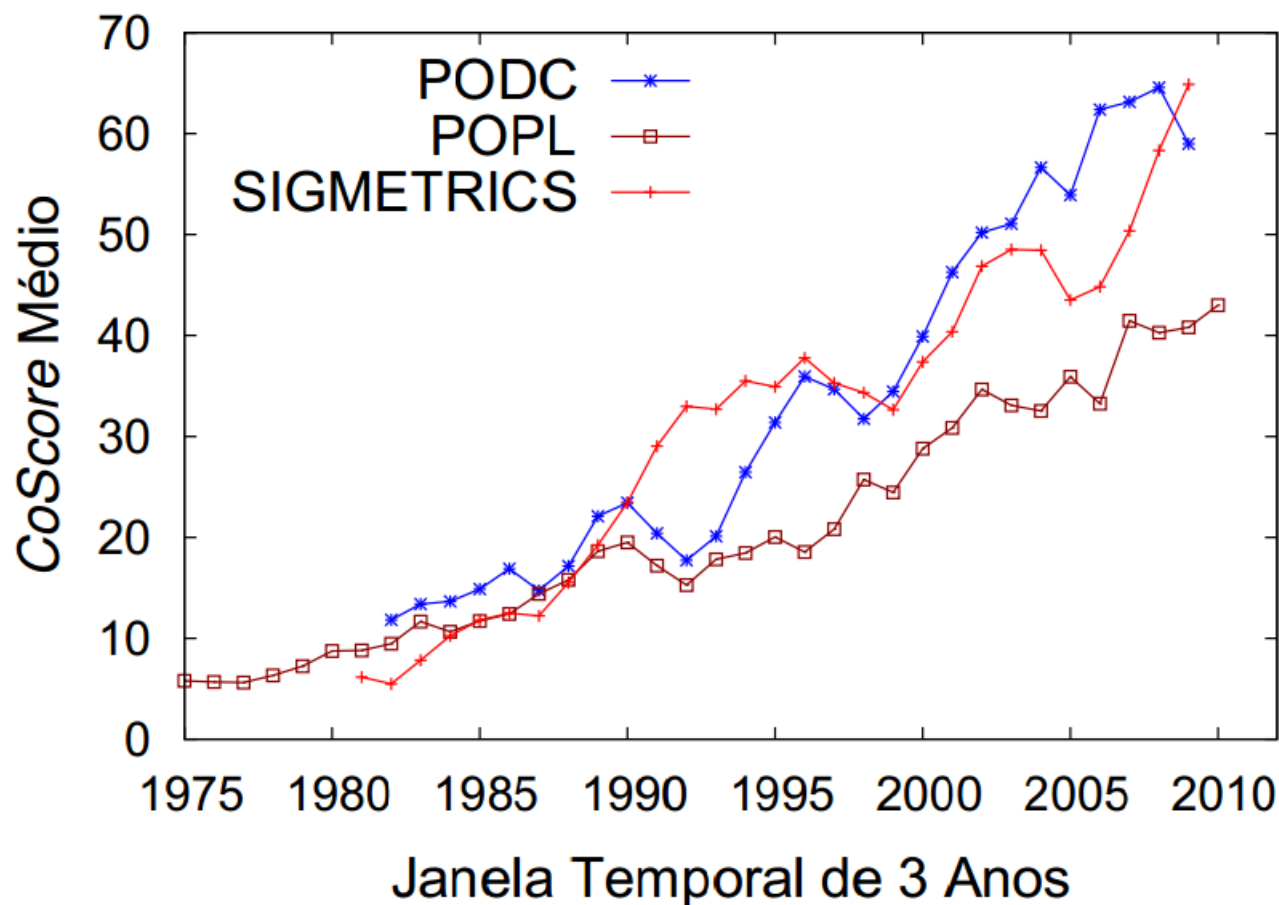
O núcleo inclui um grande número de caminhos mínimos

Influência dos Membros do Núcleo



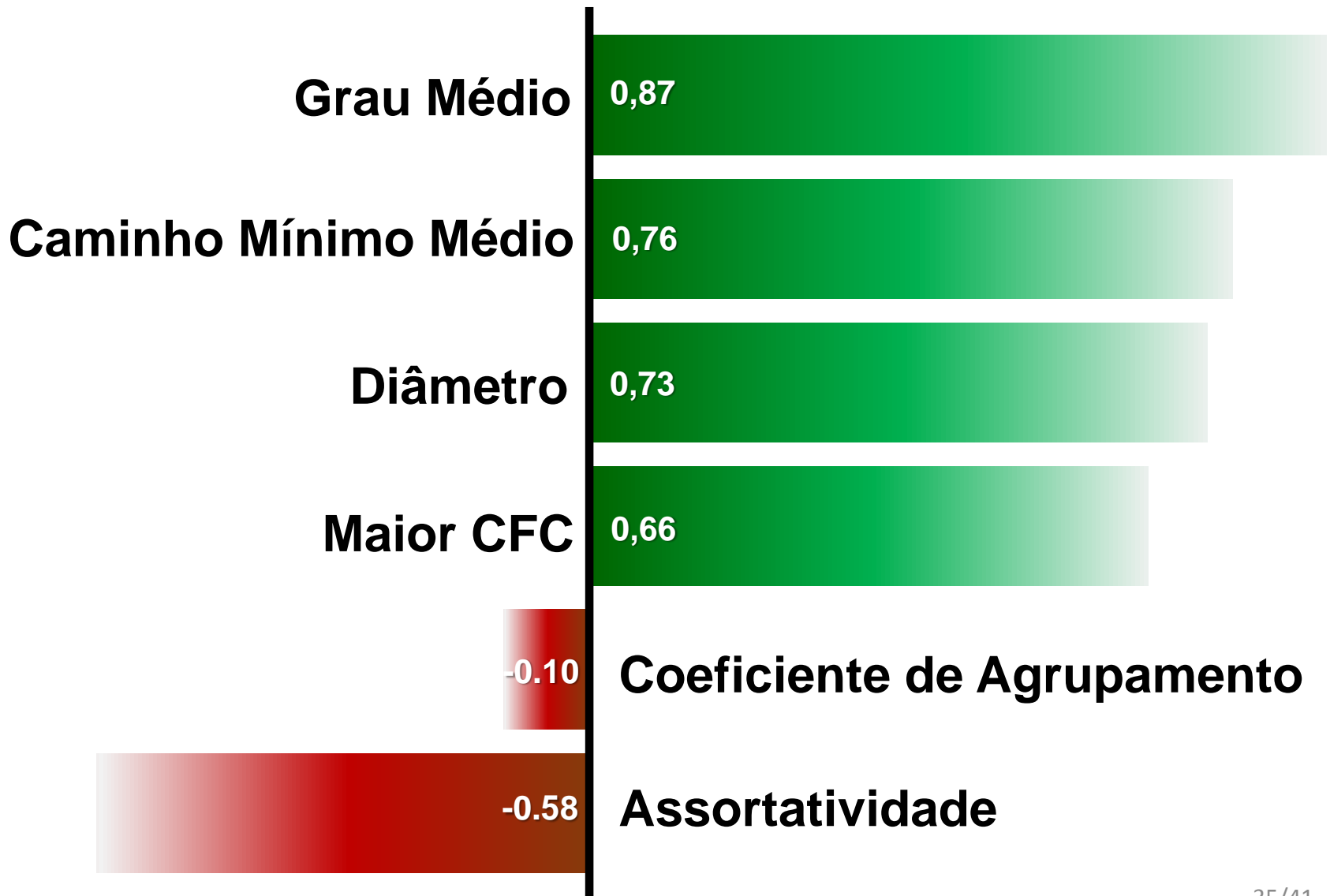
O CoScore médio de uma comunidade, em geral, aumenta ao longo do seu tempo de vida

Influência dos Membros do Núcleo

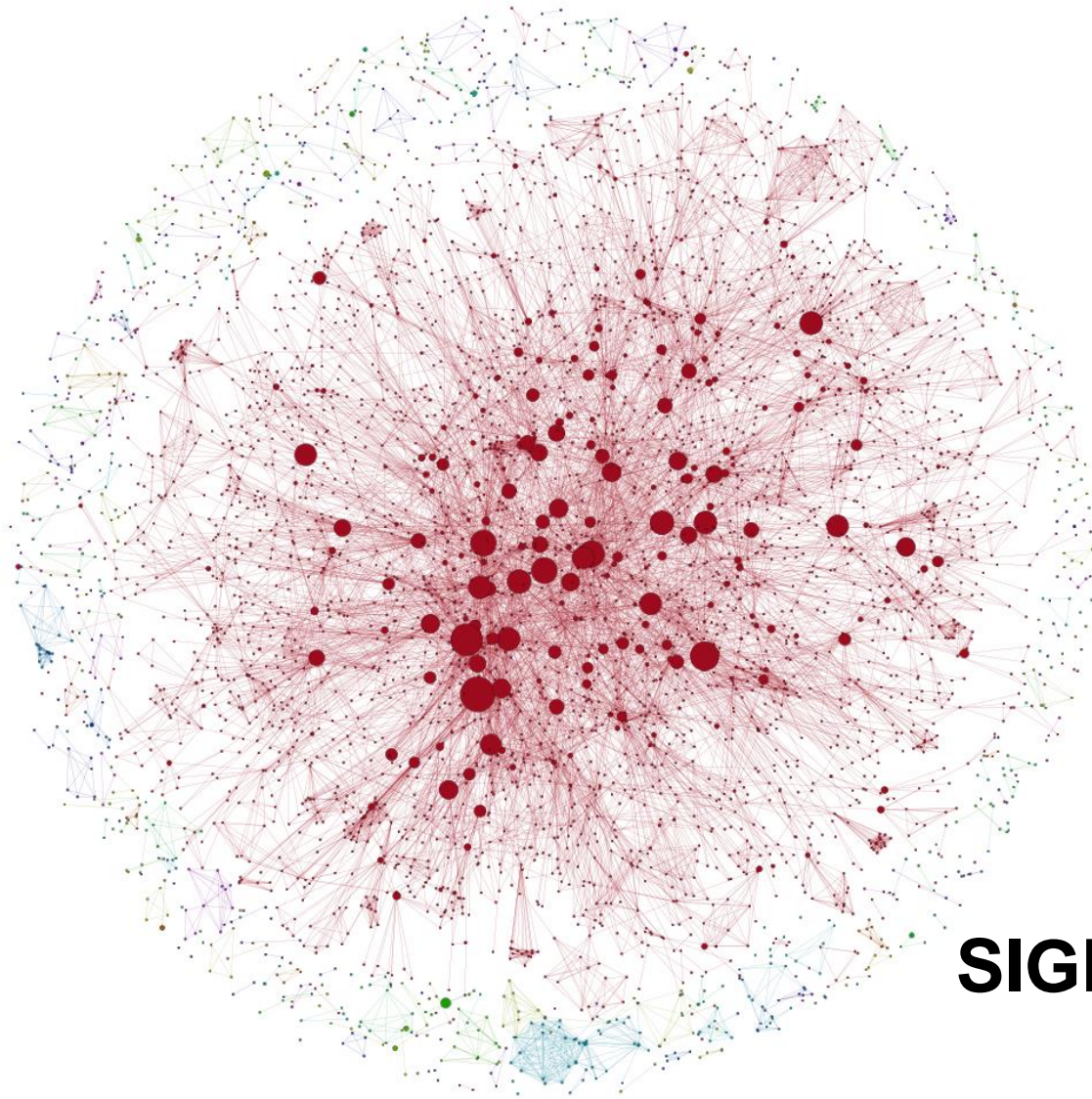


Como isso afeta a rede?

Forte Correlação com CoScore Médio



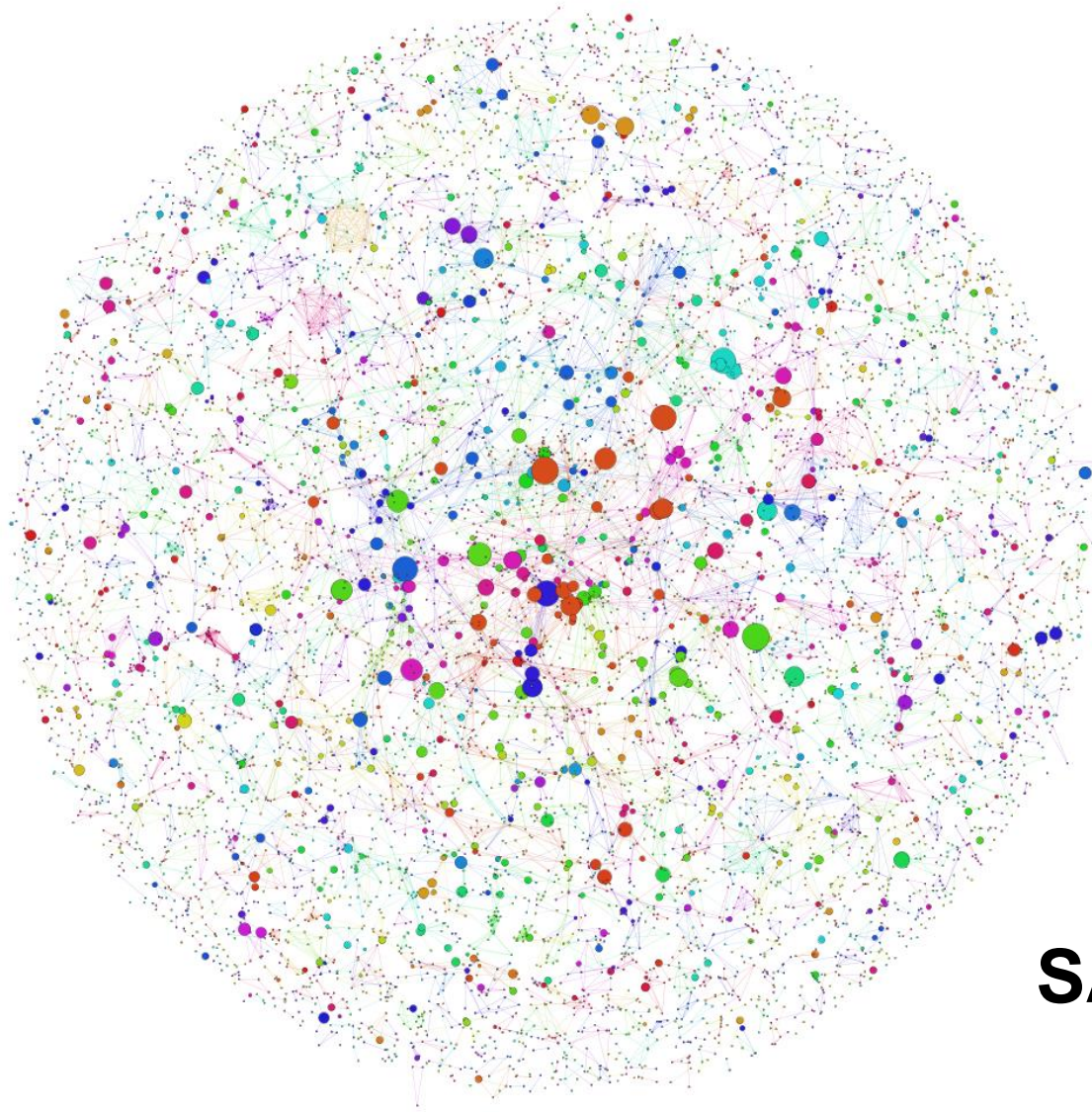
Visualização



SIGMOD

Grande número de membros do núcleo no maior CFC

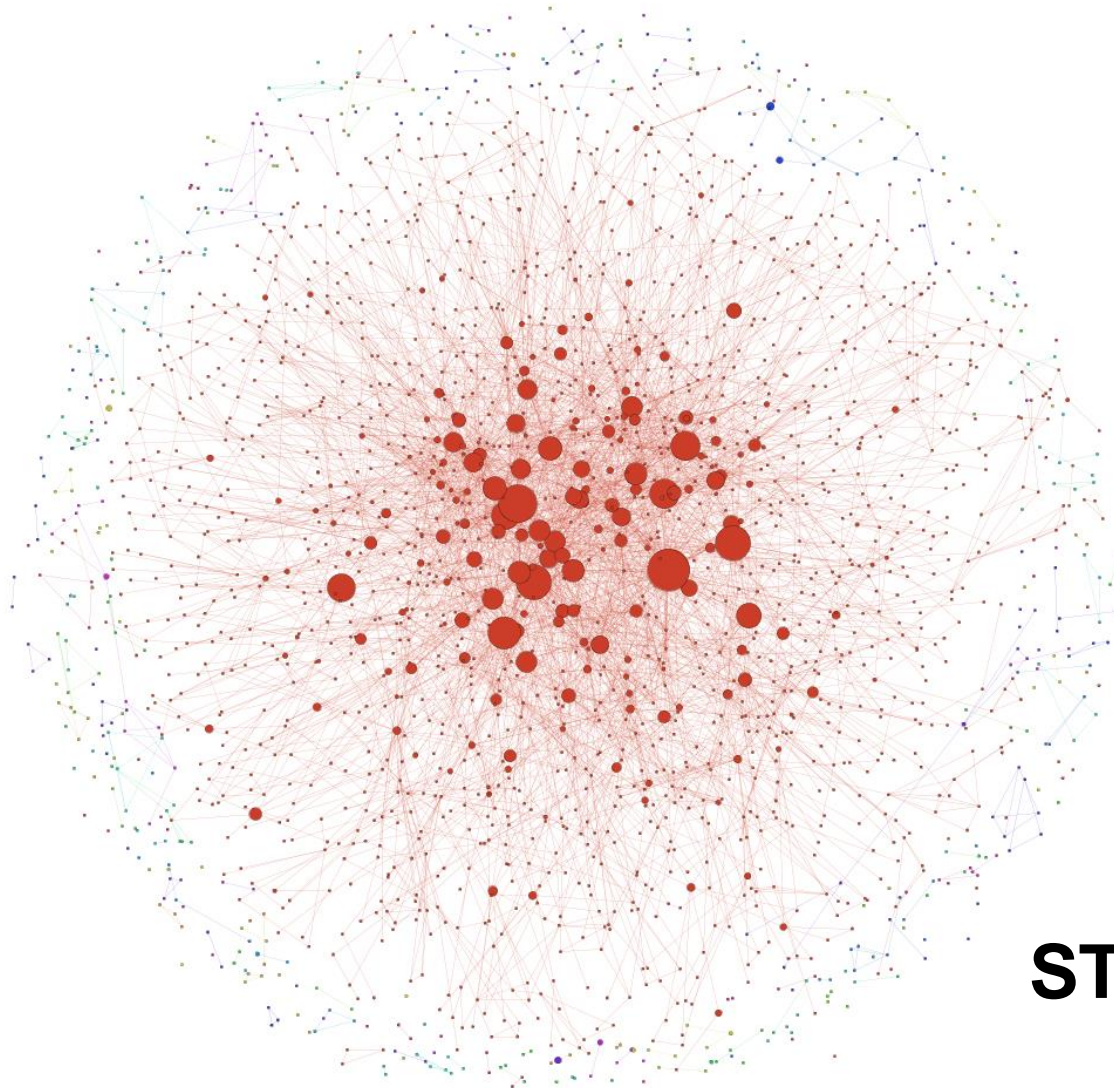
Visualização



SAC

Não possui um grande CFC bem definido

Visualização



STOC

Maior CFC muito bem definido

Conclusões

- O núcleo da comunidade:
 - É fortemente correlacionado com a variação das propriedades da rede
 - Atuam como pontes que conectam grupos
 - Aumenta o grau médio
 - Diminui a assortatividade

Conclusões

- Trabalhos Futuros:
 - Aplicação do estudo em outros contextos
 - Utilização de outras métricas de prolificidade
 - Avaliação do *CoScore* em outros contextos
 - Geração de modelos de formação de comunidades

Obrigado!

Um Estudo sobre a Evolução Temporal de Comunidades Científicas

Bruno Leite Alves

Orientador: Alberto H. F. Laender

Coorientador: Fabrício Benevenuto