

1.03.03 - Metodologia e Técnicas da Computação

VISUALIZAÇÃO E DETECÇÃO DE COMUNIDADES EM REDES COMPLEXAS

Douglas Amorim de Oliveira¹, Maria Cristina Ferreira de Oliveira²

1. Estudante do Instituto de Ciências Matemáticas e de Computação da USP
2. ICMC-USP – Departamento de Ciências de Computação / Orientadora

Resumo:

Na teoria das Redes Complexas, uma rede corresponde a um grafo que apresenta uma estrutura topológica não trivial, em termos da modelagem de informações em seus Vértices e Arestas.

A detecção de comunidades é o nome dado ao procedimento de tratar as informações de uma Rede Complexa procurando delimitar grupos de vértices altamente interligados e que tenham características em comum.

A visualização é o procedimento de tratar essas informações e exibir da forma mais legível possível para o manipulador. É o estudo da representação visual de dados, sendo informações contidas em suas arestas, vértices ou relações que podem ser abstraídas em modelo esquemático.

Nesse projeto estudou-se sobre a visualização de redes com o interesse de analisar formas de se representar interações entre 2 comunidades. Após o estudo de literatura e de tecnologias foi proposto o DAOTaVizu, uma plataforma web, que utiliza a biblioteca JavaScript D3.js para gerar visualização interativa de dados.

Palavras-chave: Detecção de Comunidades; Visualização de Dados; Redes Complexas.

Apoio financeiro: CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico.

Trabalho selecionado para a JNIC pela instituição: USP.

Introdução:

Com o advento da internet, entramos em um cenário em que a criação de dados digitais é cada vez maior, tornando assim os bancos de dados mais densos de informação. Informações estas que muitas vezes queremos processar para obter alguma característica, padrão ou procurar subgrupos. Isto tudo para pesquisas, curiosidade e até mesmo monetização, visto que muitos grandes sites e corporações usam esses dados para promover anúncios direcionados ao usuário. Por exemplo, um usuário que pesquisa ou conversa frequentemente sobre jogos digitais teria uma maior probabilidade de compra caso venha a receber anúncios de videogames e jogos do que outro usuário sem os mesmos hábitos e interesses. Com esta estratégia, as empresas digitais conseguem segmentar bem o público de interesse, fazendo assim, uma valorização desse espaço de publicidade.

É útil na pesquisa científica de mesma forma, visto que podemos usar esses dados como base, informações e características que possam vir a ser o tema alvo da pesquisa. Pode-se pegar um grande volume de dados e caracterizar grupos, a rede e filtrar por variáveis e parâmetros. E as aplicações são inúmeras, em todas as grandes áreas: nas exatas (na computação, como tráfego da internet), nas humanas (análises textuais, informações históricas, entre outras) e nas biológicas (redes neurais e sistemas de proteínas são alguns exemplos).

Com isto percebe-se que essas informações podem ser muito úteis para a comunidade e então faz-se necessário que busquemos formas eficientes para o processamento destes dados para

segmentá-los - detecção - e para apresentá-los - visualização.

O objetivo é o estudo de algoritmos de detecção de comunidades em redes, bem como o seu uso em soluções de visualização aplicáveis a grafos grandes. Essas soluções devem ser implementados em ambientes que oferecem plataformas abertas, ou utilizando bibliotecas orientadas ao desenvolvimento de visualizações interativas. Em seguida, pretende-se propor e implementar uma solução de visualização que favoreça a análise exploratória de tais redes que tenha por base a exibição e navegação em comunidades detectadas na rede.

Metodologia:

Durante o desenvolvimento do projeto, as seguintes etapas/atividades foram realizadas:

Estudo da literatura em Visualização e em Detecção de Comunidades em redes. Esta etapa foi de grande importância para a contextualização e familiarização com os temas, suas teorias e suas tecnologias, visto que foi o primeiro contato com as áreas;

Estudo e análise de plataformas/ferramentas para a visualização e análise dos dados. Inicialmente o foco principal foi a plataforma Gephi [1], que conta com uma grande variedade de recursos, tanto para a Visualização de grafos, dispondo de uma gama atrativa de layouts (Como Force Atlas, Circular Layout e Radial Axis Layout), como também para a segmentação das redes, tendo implementado algoritmos interessantes, como Modularity [2], que permite a divisão da rede em comunidades (detecção);

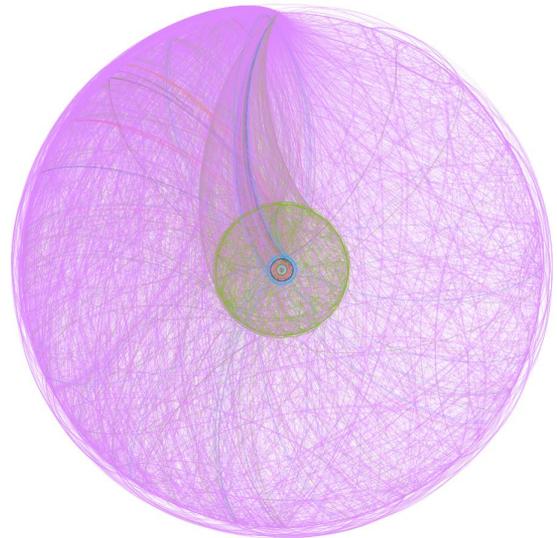


Figura 1 - Visualização da rede "High Energy Physics - Theory collaboration network" [3], gerada no Gephi, utilizando múltiplas camadas de Circular Layout

Estudo da biblioteca D3.js [4], utilizada para o desenvolvimento de Visualização de Dados acessíveis a partir de navegadores, e portanto disponibilizadas na internet;

Escolha da ferramenta na qual seria utilizada para o desenvolvimento da solução para uma exploração interativa de redes de dados. Foram analisadas as opções estudadas (a plataforma Gephi, que opera em Windows, Linux e Mac OS e o D3.js que é utilizado para o desenvolvimento de soluções de plataformas em ambiente Web). Foram utilizados para a tomada de decisão tanto os recursos como o acesso à ferramenta;

Com a discussão na etapa anterior, foi proposta uma plataforma que atendesse os objetivos de disponibilizar uma ferramenta que permita a análise exploratória dos relacionamentos entre membros de comunidades distintas;

O principal ponto considerado na decisão foi a facilidade de acesso para o usuário final. Apesar de o Gephi ser um software open-source e gratuito, o usuário que tivesse interesse em usar nossa plataforma/plugin deveria instalar o Gephi em seu computador e aprender como o mesmo funciona. Apesar de seu aprendizado ser relativamente suave, a disponibilidade de uma plataforma web é mais interessante para os

potenciais usuários, visto que a qualquer momento e de qualquer lugar o mesmo pode realizar suas visualizações apenas carregando seu arquivo de dados com a formatação, sem necessidade de instalações (e consequentemente o uso de recursos computacionais) e também sem a necessidade de aprendizado prévio de uso;

Desenvolvimento da ferramenta DAOTaVizu, utilizando os conhecimentos adquiridos nas etapas anteriores do projeto.

Resultados e Discussão:

Como resultados desta pesquisa temos a versão alfa da plataforma DAOTaVizu (disponível em: daotavizu.herokuapp.com). DAOTaVizu é uma ferramenta que tem como objetivo aprimorar métodos para análise de relacionamentos intercomunidades. Ferramenta esta que terá código aberto e será disponível para todo e qualquer usuário que tenha acesso a internet.

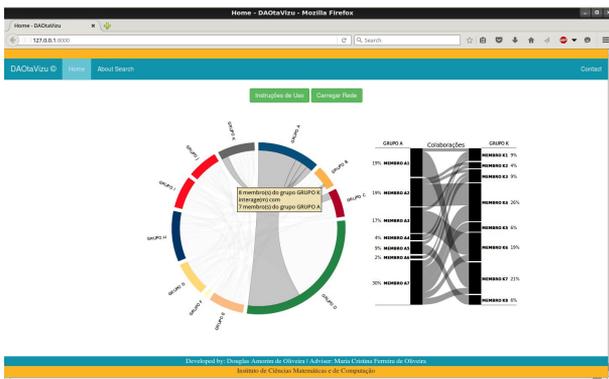


Figura 2: Visualização de uma rede fictícia exibindo as relações entre duas comunidades selecionadas, à esquerda, e a relações entre os membros dessas, à direita.

Para futuras melhorias na plataforma DAOTaVizu temos a intenção de adicionar novas funcionalidades como (i) a exploração de informações dentro das próprias comunidades, como a interação de seus componentes, (ii) detalhar informações entre as interações de membros.

Conclusões:

O problema de visualizar redes complexas é de difícil tratamento. Em geral, é preciso selecionar que tipo de informação deseja-se exibir, e escolher uma representação visual adequada. Neste projeto foi feito um estudo desse problema e desenvolvida uma solução que prioriza a exibição de comunidades e suas relações.

Referências bibliográficas

- [1] Gephi-0.9.1 - <https://gephi.org/>
- [2] github.com/gephi/gephi/wiki/Modularity
- [3] snap.stanford.edu/data/ca-HepTh.html
- [4] Bostock, M.; Ogievetsky, V.; Heer, J.; “D³ : Data-Driven Documents” (<https://d3js.org>)