

Identificação de Presença de Grupos em Grafos por Níveis de Distribuição de Grau de Centralidade

Dalton Lopes Martins
Universidade de São Paulo (USP), São Paulo, Brasil
Email: daltonmartins@usp.br

Resumo — O artigo apresenta como podemos utilizar a distribuição de grau de centralidade de uma rede social para facilitar a identificação de grupos. Baseamos nossa pesquisa em dois conceitos de base da teoria de análise de redes: o coeficiente local de agrupamento e a homofilia de grau. Os dados coletados e discutidos são apresentados com base em um experimento que analisa uma rede social de monitores e tutores participantes de um curso de formação à distância. Os resultados indicam que a distribuição de grau de centralidade pode ser usada como base na identificação dos maiores e mais conectados grupos presentes numa rede.

Palavras chave—grafos, redes sociais, grau de centralidade, comunidades, grupos.

I. INTRODUÇÃO

O uso de grafos para modelar e visualizar relações de nós em uma configuração de rede, seja de hiperlinks em páginas na Internet, troca de e-mails entre pessoas, redes sociais *online* ou mesmo as configurações da relação de neurônios no cérebro humano tem sido uma tendência na análise de problemas computacionais na última década. No entanto, grafos podem apresentar padrões complexos de difícil percepção visual, exigindo, portanto, formas de agrupamento e redução de objetos visualizados como modo de facilitar a identificação de tendências e padrões. A visibilidade a esses padrões e possibilidades de análise que lhe são decorrentes é uma das questões centrais da pesquisa sobre análise de redes sociais, dado que esta é intuitivamente baseada na noção de que os eventuais padrões de conectividade entre os atores que compõem uma rede têm importantes consequências para esses mesmos atores [2].

Há diversas formas de identificar e mapear os padrões mais importantes de uma rede. Importantes questões de pesquisa para as áreas da comunicação e informação têm sido respondidas a partir dessa análise desse mapeamento de padrões: quais os atores em uma rede que parecem pautar suas discussões? Qual importante uma página web pode ser considerada? Quais genes são letais se eliminados de uma cadeia de conexões? Quão crítico é um roteador para o fluxo de informações de uma determinada rede?

Um dos elementos fundamentais de pesquisa para esses problemas apresentados é a ideia de centralidade de um nó

em relação aos demais, sendo a forma mais simples de centralidade o número de conexões diretas que um nó possui em seu entorno [1].

Os nós possuem número de conexões que em redes reais não apresentam distribuições homogêneas, seguindo, em geral, padrões de distribuição regulados por uma lei de potência, indicando poucos nós com muitas conexões e muitos nós com poucas conexões. A distribuição das conexões também apresenta um padrão que não é global, ou seja, há grandes concentrações para um grupo específico de nós e poucas para outros grupos [7].

A descoberta desses padrões de distribuição pode levar a identificação de grupos de nós, permitindo maior capacidade de síntese, novos modos de visualização e exploração dos recursos disponíveis nas conexões em uma rede [5]. Outro modo de produzir síntese melhorando os modos de visualização de padrões em redes é remover as conexões mais fracas do grafo, reduzindo o número de informações disponíveis, facilitando a identificação das relações e posições estruturais dos nós mais centrais na rede [6]. A ideia por trás dessas duas formas de tratamento dos grafos de rede é que identificando os atores mais centrais identifica-se onde a ação de articulação mais importante está acontecendo no contexto relacional dos nós em questão [8].

Observa-se que o ponto central desses dois modos de identificar grupos e simplificar a complexidade informacional dos grafos consiste na análise de como se distribuem as conexões entre os nós, seja identificando as mais intensas e recorrentes analisando-as em separado, seja identificando as mais fracas e periféricas para poder retirá-las do grafo, simplificando sua análise.

Uma comunidade em um grafo consiste de um grupo de nós que estão relativamente mais conectados entre si, ou seja, com maior densidade de conexão do que com os outros nós pertencentes ao grafo [3]. Dois fundamentos oriundos da teoria de análise de redes podem ser utilizados em conjunto para explicar como é possível observar a distribuição das conexões dos nós em uma rede visando identificar agrupamentos que possam ser considerados comunidades, a saber, o coeficiente de agrupamento local e a homofilia de grau.

É apresentado um experimento neste trabalho indicando que a análise da distribuição do grau de centralidade dos nós de uma rede, ou seja, a medida que indica o número de conexões diretas que um nó possui, pode ser utilizada para a

identificação automática de níveis de conexões dos nós em um grafo, facilitando a identificação da presença de grupos. Evidências em [6] e [11] reforçam o uso desta técnica como uma forma de identificação de grupos em grafos de redes sociais.

A questão que se coloca neste trabalho é se o uso da identificação de grupos em grafos de redes sociais através da técnica de remoção das conexões mais fracas poderia ser utilizada como uma das técnicas de monitoramento das redes no apoio a identificação padrões relevantes.

II. COEFICIENTE DE AGRUPAMENTO LOCAL

O coeficiente de agrupamento local é calculado para um nó i através da relação entre o número de pares de vizinhos de i que estão conectados entre si e o número de pares de vizinhos de i . Um resultado importante no contexto deste trabalho é o fato de quando calculado o coeficiente para todos os vértices de uma rede surge um padrão: na média, os nós com maior grau de conexão tendem a ter um menor coeficiente de agrupamento local [4].

Uma possível explicação para esse resultado é o fato de que grupos de nós tendem a estar mais conectados entre si, levando a que grupos pequenos tenham menor grau de conexão por estarem conectados com menos nós e grupos maiores apresentem maior grau de conexão por estarem conectados com mais nós. Ao mesmo tempo, o coeficiente de agrupamento local tende a ser maior em grupos menores, pois haverá proporcionalmente mais vizinhos do nó i que estarão conectados entre si do que em grupos maiores [4].

Logo, é possível inferir dos resultados acima apresentados que grupos de nós podem ser considerados grupos pequenos ou grandes com base no número de nós que fazem parte desses grupos, o que impacta diretamente no grau de conexão que é também compartilhado, em média, dentro desses grupos. Grupos menores tendem a possuir um grau de conexão menor entre si, sendo que grupos maiores tendem a possuir um grau de conexão maior entre si.

Apresenta-se agora um outro conceito oriundo da teoria de análise de redes que permite complementar as tendências observadas em relação ao coeficiente local de agrupamento.

III. HOMOFILIA DE GRAU

A homofilia é a tendência de que nós que se conectam entre si são semelhantes de algum modo. De particular interesse para este trabalho é a homofilia de grau, que indica a tendência de que os nós que se conectam entre si tendem a possuir graus de conexão que são similares [4].

Experimentos indicam que a homofilia de grau não é identificada em qualquer tipo de rede, sendo que há claros indícios de que ela é mais evidente em redes sociais, onde os nós representam pessoas. Uma possível explicação para isso consiste na observação de que as redes sociais tendem a ser mais divididas em grupos de nós do que outros tipos de redes. Se uma rede é dividida entre grupos onde os nós estão mais conectados entre si do que outros, pertencentes a outros grupos mais isolados, novamente, pode-se dizer que nós em pequenos grupos tendem a possuir menor grau de conexão do que nós pertencentes a grupos maiores [4].

IV. DISTRIBUIÇÃO DO GRAU DE CENTRALIDADE E A IDENTIFICAÇÃO DE GRUPOS DE NÓS

Pode-se relacionar, no caso específico das redes sociais, os resultados apresentados do coeficiente de agrupamento local e da homofilia de grau para analisá-los em conjunto, considerando seus efeitos na identificação de grupos de nós em uma rede.

Considerando que o objetivo deste trabalho seja identificar a presença dos principais grupos de pessoas que atuam em conjunto em uma rede qualquer, se infere dois aspectos desses grupos que serão utilizados a seguir: os maiores grupos, além de serem compostos por mais nós do que os grupos menores, são aqueles onde os nós possuem maior grau de centralidade.

Levando isso em consideração na análise, se utiliza apenas a curva de distribuição de centralidade dos nós de uma rede como elemento orientador na determinação de níveis de centralidade que podem facilitar a encontrar grupos de nós. Na prática, se propõe neste experimento utilizar a curva de distribuição de centralidade para identificar determinados níveis de grau de centralidade e retirar nós que estejam abaixo desses níveis do grafo, observando se a organização dos nós restantes apresenta claros indícios de comunidades distintas.

A aplicação desse tipo de recurso para analisar um grafo poderia ser implementada de modo automático e configurável em algum possível software de análise de redes, permitindo que pesquisadores de diferentes áreas do conhecimento, sobretudo aqueles oriundos de áreas que dialogam com as ciências sociais e humanas, menos afeitos com as complexidades matemáticas e operacionais dos grafos, pudessem utilizar uma ferramenta de fácil acesso na busca e identificação de comunidades nas redes sociais com as quais estejam trabalhando.

V. METODOLOGIA

Utilizou-se para a realização deste trabalho experimental um banco de dados de um sistema de ensino à distância, utilizado por aproximadamente 1.400 alunos de um curso de formação para inclusão digital em todo o território brasileiro. O curso é promovido pelo Programa Nacional de Apoio à Inclusão Digital nas Comunidades – Telecentros.BR, que é uma ação do Governo Federal de apoio à implantação de novos espaços públicos e comunitários de inclusão digital e o fortalecimento dos que já estão em funcionamento em todo o território [9]. Uma ação de apoio do programa Telecentros.BR é a Rede Nacional de Formação para Inclusão Digital, instância responsável pela construção, manutenção e aplicação de um curso de formação para os jovens bolsistas monitores dos telecentros. Logo, é a partir dessa ação pública que coletou-se os dados de amostra deste trabalho. Os dados foram obtidos diretamente do banco de dados do sistema de apoio a promoção do curso.

Dentro do ambiente de ensino à distância, foi disponibilizado pela Rede de Formação um espaço para conversação instantânea entre os usuários que estiverem conectados ao mesmo tempo na plataforma. Esse dispositivo apresenta uma lista de usuários, sendo que quando algum deles desejasse conversar com outro, poderia apenas clicar no

seu nome e uma janela de conversação seria disponibilizada para troca de mensagens. As figuras 1 e 2 mostram como esse dispositivo opera.



Figura 1. Exemplo de visualização da lista de usuários conectados.

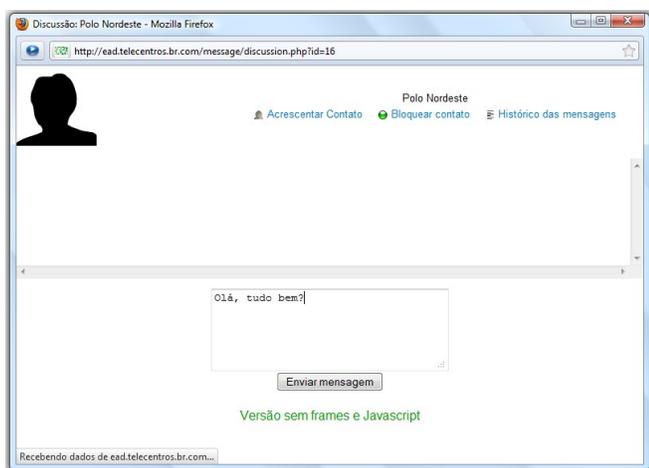


Figura 2. Exemplo de janela de conversação entre usuários.

Para a modelagem da rede, consideramos apenas o módulo de troca de mensagens instantânea entre os alunos. Essa escolha se deve a 2 fatores principais: o módulo fica visível o tempo todo para os alunos, indicando quem está *online* e disponível para conversar em qualquer momento do curso; o módulo é utilizado de forma espontânea, ou seja, não há nenhuma atividade do curso que promova ou incentive seu uso específico. Além dos fatores técnica acima mencionados, uma questão pedagógica estimulou essa escolha: há ou não uma rede sendo formada de modo espontâneo pelos usuários? Se há, de que modo essa rede poderia ser mais um espaço onde ações explícitas de formação poderiam ocorrer?

Os dados coletados para a realização deste estudo representam 6 meses de interação entre os participantes. Os alunos foram modelados em cor verde e os tutores, responsáveis pela aplicação da formação, em cor amarela. É importante ressaltar que na organização do curso de formação

em análise, os tutores eram responsáveis diretamente por um grupo de até 30 monitores, realizando seu acompanhamento diário de atividades, dúvidas e eventuais comunicações que se fizessem necessárias. Foram consideradas todas as mensagens trocadas pelos participantes nesse período, totalizando 15712 mensagens trocadas entre os mesmos.

Após a extração dos dados, foi construída uma matriz de correspondência entre usuários, identificando cada participante por uma chave única e exclusiva quando de seu cadastro na plataforma de ensino. A matriz indica todas as relações que foram estabelecidas entre usuários, permitindo que se calculasse o grau de centralidade não apenas pelo fato dos usuários terem ou não se conectado entre si alguma vez, mas também levando em consideração no cálculo a quantidade de vezes que os usuários interagiram.

Desse modo, o número de conexões que um usuário estabelece na rede está diretamente conectado a dois fatores: com quantas pessoas esse usuário se conectou e com que intensidade ele manteve essas conexões.

Nesta rede específica, isso permitiu analisar o efeito que essa intensidade de conexões causa no grau de centralidade, sobretudo considerando que a rede é composta de perfis de usuários distintos, onde se têm monitores que são os alunos que participam do curso de formação, mas também os tutores que, em tese, deveriam se conectar com mais usuários na plataforma, pois são eles os responsáveis diretos por seus grupos de alunos.

O grau de centralidade e a visualização dos sociogramas apresentados a seguir foi realizada com o software Pajek [10].

VI. RESULTADOS

O grau de centralidade dos monitores e tutores foi calculado e sua distribuição pode ser visualizada na figura 3.

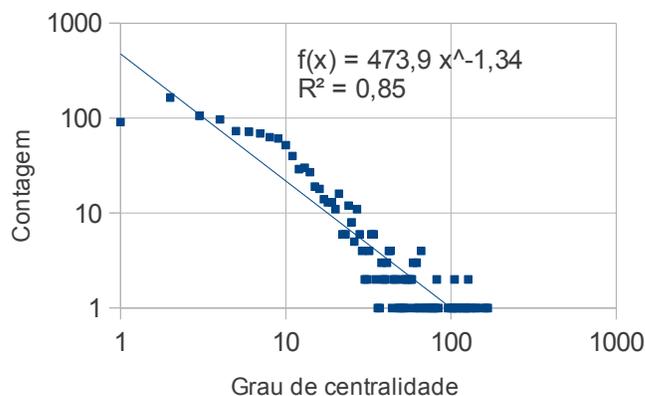


Figura 3. Distribuição dos graus de centralidade

Calculou-se a regressão geométrica da distribuição, cujo resultado é apresentado na figura 3. O grau máximo de centralidade encontrado na rede foi de 167, sendo que apenas um nó obteve esse valor, representado pelo ponto mais a direita no eixo x da figura 3. O grau mínimo de centralidade foi de 0, sendo que 62 nós apresentam esse estado. Como o gráfico da figura 3 foi normalizado em escala log nos eixos x e y esses nós não são apresentados na figura. O maior número

de nós que compartilha o mesmo grau foi de 165, representados pelo pico do eixo y na figura 3, sendo que estes nós possuíam grau de centralidade 2.

Com base no resultado da distribuição e explorando visualmente as imagens de redes formadas a partir de diferentes limiares de grau de centralidade como filtro dos nós participantes, foram identificados três níveis de centralidade que poderiam gradativamente reduzir a presença dos nós com menor número de conexões, facilitando a identificação da formação de grupos. Apresenta-se na tabela I, a seguir, como essa distribuição se dá em termos do número de nós em cada uma e o quanto representam do total de nós na rede.

TABELA I
Níveis de Centralidade

Grau de centralidade	Número de nós	% do total de nós
10	447	34,23%
15	269	20,60%
20	192	14,70%

O corte da rede pelo grau de centralidade 10 resulta em aproximadamente um terço da rede que ainda se mantém conectada, já no grau 15 esse número é reduzido para um quinto e para o grau 20 é reduzido para aproximadamente um sexto.

Os níveis de corte propostos permitiram gerar sociogramas, imagens das conexões nos grafos, que são apresentados nas figuras 4, 5 e 6.

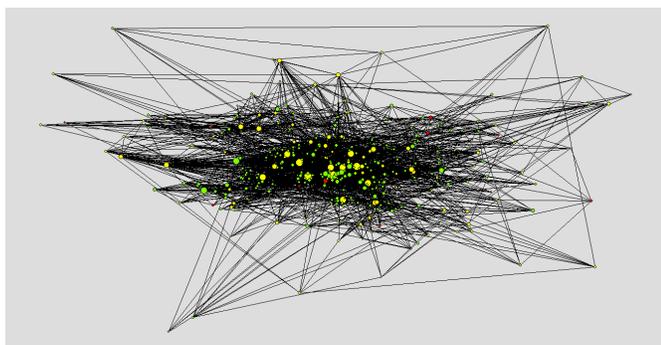


Figura 4. Nós com mais de 10 conexões

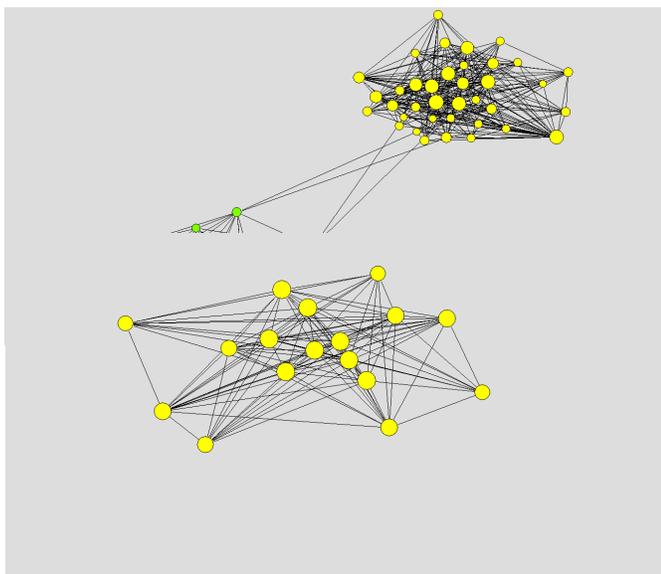


Figura 5. Nós com mais de 15 conexões

Figura 6. Nós com mais de 20 conexões

Os grafos das figuras 5 e 6 apresentam padrões de conexão que demonstram visualmente a redução de complexidade visual apresentada na figura 4, já permitindo identificar diferentes padrões organizacionais no modo como os nós estão conectados entre si.

VII. DISCUSSÃO

Os resultados apresentados na seção anterior permitem avaliar o efeito gerado na forma e estrutura das redes pelo uso da curva de distribuição de grau de centralidade como elemento experimental para a identificação de grupos em grafos de redes sociais.

Uma primeira observação que é importante ressaltar é que o uso de níveis de grau de centralidade para reduzir o número de nós presente na rede parece não apenas reduzir os elementos presentes na rede, mas ressalta níveis de organização e conectividade entre os nós que indica formações de agrupamento de nós bastante evidentes, como se observa nas figuras 5 e 6.

Na figura 4, se observa uma rede onde não se identifica com clareza a presença de algum grupo específico, dado que apenas se visualiza alguns nós periféricos e um núcleo altamente concentrado. São evidentes alguns pontos amarelos e verdes, demonstrando os tutores e monitores que apresentam maior grau de centralidade. No entanto, ainda não é possível determinar pela análise do sociograma como se dá a conexão entre esses usuários e se elas formam algum de tipo de grupo específico.

Na figura 5, se observa como o corte no grafo pelo grau de centralidade 15 deixa evidente duas comunidades bastante distintas e com bem poucas conexões entre elas. Os monitores, representados pela cor verde, possuem grau de conectividade menor, o que é indicado pelo tamanho dos nós na imagem, porém apresentam conexão muito mais intensamente com outros monitores do que com seus próprios tutores, representados pela cor amarela, a exceção de poucas conexões que interligam os dois grupos. Identifica-se também na figura 5 apenas um tutor que está bem mais próximo do grupo de monitores do que do grupo de tutores. Esse tutor se conecta mais intensamente com outros monitores do que com outros tutores.

Na figura 6, se observa que o corte pelo grau de centralidade 20 evidencia a comunidade que se conecta mais intensamente entre si na rede em questão, a comunidade de tutores. Como se observa na imagem, o tamanho dos nós é similar, indicando que todos os presentes apresentam um grau de centralidade bastante próximo. Não se identifica a

presença de nenhum monitor nessa configuração, indicando que os tutores de fato se falam mais intensamente entre si do que com os próprios monitores que acompanham, o que se evidencia através da análise do grafo de sua formação de grupo.

Os tutores e monitores, apesar de trabalharem em conjunto para o desenvolvimento do curso de formação, se configuram como duas comunidades que apresentam características distintas. Ao que tudo indica, o papel e a posição que exercem no decorrer de seu curso de formação determina mais fortemente sua configuração e pertencimento a um grupo característico do que o fato de os tutores estarem mais em contato com um grupo específico de monitores que acompanham.

Algumas hipóteses podem ser levantadas em relação a esse comportamento observado. A primeira delas é que os tutores, por compartilharem o mesmo papel e realidade de trabalho, trocam mensagens entre si como um modo de compartilharem estratégias, trocaram impressões, conteúdos, exercícios e modos de operar que lhes são característicos. A segunda é que os tutores trabalham para organizações específicas, logo têm uma maior chance de se conhecerem pessoalmente e atuarem fisicamente seja no mesmo espaço ou de algum modo próximos, facilitando com que possam estabelecer vínculos pessoais de outra ordem que podem afetar o nível de conversação no ambiente de formação. A terceira é que os monitores podem identificar na listagem de usuários conectados, apresentada na figura 1, quais deles são tutores e monitores, facilitando com que possam utilizar estratégias diversas quando buscam trocar mensagens, permitindo com que encontrem outros monitores para compartilhar ideias, dificuldades, estratégias de atendimento e assuntos em comum que preferem trocar com monitores do que com seus próprios tutores.

VIII. CONCLUSÃO

O uso de níveis de centralidade a partir da análise da distribuição de seus graus parece indicar um procedimento que reduz a complexidade da informação em grafos de redes, facilitando a identificação da presença de grupos e análise de padrões.

As evidências coletadas e apresentadas nesse trabalho parecem indicar que o uso dos resultados teóricos obtido pela homofilia de grau e o coeficiente local de agrupamento podem ser utilizados para construir dispositivos baseados na distribuição do grau de centralidade para identificar grupos de nós em grafos de redes sociais. A simplicidade de aplicação desse recurso pode ser incorporada e manipulada em softwares de análises de redes por usuários que tenham pouco conhecimento da teoria dos grafos e análise de redes.

A distribuição do grau de centralidade parece ser favorável na identificação dos maiores e principais grupos de nós conectados entre si. Não foram encontradas evidências que permitam utilizar esse método para a identificação de pequenos grupos, que possam representar ilhas de nós conectados entre si ou novos grupos ainda em estágio de formação.

AGRADECIMENTOS

Gostaríamos de agradecer a Rede Nacional de Formação para Inclusão Digital, as instituições que fazem parte da Rede, a saber: Universidade Federal do Pará, Universidade Federal de Minas Gerais, Universidade do Estado da Bahia, ONG Coletivo Digital, ONG Programando Futuro, ONG Maristas, ONG Idear. Em especial, gostaríamos de agradecer a coordenação executiva da Rede de Formação e do programa Telecentros.BR, alocada na Secretaria de Inclusão Digital do Ministério das Comunicações, pelo apoio na realização deste trabalho.

REFERÊNCIAS

- [1] E. D. Kolaczyk. Statistical analysis of network data: methods and models. Springer. 1st ed. 2009. p. 386.
- [2] L. Freeman. The development of social network analysis: a study in the sociology of science. Empirical Press. 1st ed. 2004. p. 205.
- [3] M. A. Porter, J-P Onnela, P. J. Mucha. Communities in Networks. Notices of the American Mathematical Society, Vol. 56, No. 9: 1082-1097, 1164-1166, 2009
- [4] M. E. J. Newman. Networks: an introduction. Oxford University Press. 1st ed. 2010. p. 772.
- [5] M. Girvan, M. E. J. Newman. Community structure in social and biological networks. *PNAS.*, 2002. vol. 99 no. 12: 7821-7826.
- [6] O. Persson. Short, strong and simple mapping of research fields. *Proceedings of ISSI 2011* - Durban, July 4-7, 2011.
- [7] S. Fortunato. Community detection in graph. *Phys. Rep.*, 2010. 486: 75-174.
- [8] S. Wasserman, K. Faust. Social network analysis: methods and applications. Cambridge University Press. 18th ed. 2009. p. 825.
- [9] Telecentros.Br – Documento orientador do programa. Disponível em: http://www.inclusaodigital.gov.br/telecentros/documentos/documento_positivo.pdf. Acessado em 05/02/2012.
- [10] V. Batagelj, A. Mrvar. PAJEK - Reference Manual: list of commands with short explanation - version 2.05. Disponível em: <http://pajek.imfm.si/lib/exe/fetch.php?media=dl:pajekman205.pdf>. Acessado em 05/02/2012.
- [11] W. de Nooy, A. Mrvar, V. Batagelj. Exploratory social network analysis with Pajek. Cambridge University Press. 1st ed. 2005. p. 334.