

## Uma descrição natural das primeiras mil mensagens públicas da Rede Metareciclagem

Renato Fabbri,<sup>1, a)</sup> Vilson Vieira,<sup>1, b)</sup> Caleb Mascarenhas Luporini,<sup>2, c)</sup> Ricardo Fabbri,<sup>3, d)</sup> Glerm Soares,<sup>4, e)</sup> Maira Begalli,<sup>5, f)</sup> Felipe Fonseca,<sup>5, g)</sup> Jader Gama,<sup>6, h)</sup> and Lista Metareciclagem<sup>7, i)</sup>

<sup>1)</sup> Instituto de Física de São Carlos, Universidade de São Paulo (IFSC/USP)

<sup>2)</sup> LabMacambira.sourceforge.net

<sup>3)</sup> Instituto Politécnico do Rio de Janeiro, Universidade do Estado do Rio de Janeiro (IPRJ/UERJ)

<sup>4)</sup> Movimento dos Sem Satélites (MSST)

<sup>5)</sup> Rede Metareciclagem

<sup>6)</sup> Coletivo Puraque

<sup>7)</sup> Sociedade civil

(Data: 23 de Janeiro de 2013)

Este trabalho apresenta uma primeira observação das redes de respostas das primeiras mil mensagens públicas da lista de emails da Rede Metareciclagem. O propósito é a verificação de estruturas naturais e uma reflexão sobre as propriedades e possibilidades organizacionais que delas provêm. Algumas das características são típicas de fenômenos sociais, como as distribuições livres de escala e a propriedade de pequeno mundo. Estas sugerem que a rede utilizada capture aspectos fundamentais do fenômeno estudado, embora seja bastante simples. Já outras propriedades verificadas, em especial as assimetrias das relações e as proporções entre graus e forças, parecem ser inéditas e sugerem utilidade do aprofundamento com outras amostras de mensagens. Este trabalho acionou as comunidades já nos primeiros momentos, ampliando as colaborações para um amadurecimento público das questões e validade das discussões e resultados. A análise realizada tem potenciais implicações no entendimento de mecanismos sociais fundamentais, bem como levanta tópicos de grande relevância na era da internet massificada, como questões de ética e publicidade dos dados, e mesmo de autoconsciência de mecanismos e atitudes sociais visíveis em larga escala. Este trabalho também está sendo amadurecido para o projeto de meios eficientes de articulação e para facilitar fluxos de informação.

PACS numbers: 89.75.Fb, 05.65.+b, 89.65.Ef

Keywords: social network analysis, email networks, pattern recognition, statistics

### I. INTRODUÇÃO

Atividades de comunicação e coordenação são centrais em desenvolvimentos tecnológicos e outras atividades descentralizadas territorialmente, mas são difíceis de serem observadas em contextos tradicionais (não-abertos, não-livres) por prevalecer a comunicação informal ou de acesso restrito. Comunidades de *software* livre, e outras igualmente engajadas, por outro lado, utilizam a internet como meio de comunicação e tipicamente conduzem as discussões de uma forma aberta, pública. Em particular,

todo *software* livre possui uma ou mais listas públicas onde os membros se comunicam e coordenam suas atividades. Todos os traços destas listas são arquivados e disponibilizados para consulta e estudo. Estes arquivos de emails, em conjunto com repositórios de códigos e outros artefatos disponíveis *online*, constituem recursos únicos e valiosos para estudos de comunicação e coordenação social. Em especial, os atuais arquivos públicos do GMANE contam com mais de 20 mil listas de *emails* e 130 milhões de mensagens.<sup>1</sup> Com o devido tratamento, pode-se observar redes sociais e tratar questões de interesse. Essas questões dizem respeito às intensidades de participação e suas tipificações, à distribuição de lastros e status, à relação entre atividades nos emails da lista e atividades fora da lista e às dinâmicas de fluxo de informação e opinião inerentes à topologia observada. Com isso, busca-se observar princípios gerais e fundamentais desses fenômenos, pois suas presenças indicam que a rede em mãos é representativa do fenômeno estudado, além de facilitar tipificações e comparações para caracterização minuciosa.

A seção seguinte é uma descrição do fenômeno em estudo e dos dados em foco; na sequência, uma especificação do tratamento dos dados garante a concisão do método; por fim, há uma análise inicial dos dados obtidos. O Apêndice apresenta alguns dos dados de forma direta. O IFSC/USP e NILC/USP possuem atuação em

<sup>a)</sup> <http://labmacambira.sourceforge.net>; Electronic mail: [renato.fabbri@gmail.com](mailto:renato.fabbri@gmail.com)

<sup>b)</sup> <http://automata.cc>; Electronic mail: [vilson@void.cc](mailto:vilson@void.cc)

<sup>c)</sup> <http://calebav.com>; Electronic mail: [calebml@gmail.com](mailto:calebml@gmail.com)

<sup>d)</sup> <http://www.lems.brown.edu/~rfabbri>; Electronic mail: [rfabbri@gmail.com](mailto:rfabbri@gmail.com)

<sup>e)</sup> <http://glerm.devolts.org>; Electronic mail: [orga-nismo@gmail.com](mailto:orga-nismo@gmail.com)

<sup>f)</sup> <http://mairabegalli.net>; Electronic mail: [ce0064@gmail.com](mailto:ce0064@gmail.com)

<sup>g)</sup> <http://glerm.devolts.org>; Electronic mail: [felipefonseca@gmail.com](mailto:felipefonseca@gmail.com)

<sup>h)</sup> <http://puraque.org.br/blog/?tag=jader-gama>; Electronic mail: [gama.puraque@gmail.com](mailto:gama.puraque@gmail.com)

<sup>i)</sup> <http://rede.metareciclagem.org>; Electronic mail: [metareciclagem@lists.riseup.net](mailto:metareciclagem@lists.riseup.net)

destaque nas áreas de Redes Complexas e Processamento de Linguagem Natural, ambas cruciais para este tipo de estudo.

## II. CONVERSADORES E ATUANTES: OS DADOS

De forma paradigmática, uma lista de emails de um *software* livre é um fórum público. Qualquer um pode postar mensagens na lista. As mensagens postadas são enviadas para todos os inscritos. Autores das mensagens incluem desenvolvedores, relatores de *bugs*, contribuidores que submetem *patches* mas não possuem privilégios para *commit* e usuários comuns. As listas de emails podem ser bastante ativas, na casa dos milhares de mensagens por ano, como a lista do *Apache*, ou de dezenas de milhares por ano, como a do *gcc*. Nestes casos, o número de usuários diferentes que mandaram mensagens alcança alguns milhares. Considera-se que  $\approx 73\%$  das mensagens são respostas.<sup>2</sup> Uma resposta  $b$  para uma mensagem  $a$  é uma indicação de que o autor  $s_b$  de  $b$  achou o que disse o autor  $s_a$  interessante; portanto a resposta  $b$  indica informação fluindo de  $s_a$  para  $s_b$ . Além disso, também é uma indicação de status, já que  $s_b$  achou a mensagem de  $s_a$  digna de ser lida, e digna de ser respondida. Na lista metareciclagem, encontra-se  $\approx 6000$  mensagens anuais, 79.7% delas são respostas a outras mensagens, um pouco mais do que o esperado.

O nível de atividade dos participantes varia dramaticamente. No caso das 1000 mensagens iniciais, foram encontrados 71 autores diferentes. O membro mais ativo enviou 169 das mensagens (sozinho foi responsável por 16.9% do total de mensagens enviadas), vários membros mandaram somente uma mensagem ou pouquíssimas ( $\approx 40$  pessoas mandaram menos que 4 mensagens), alguns membros são intermediários. As mensagens refletem interações de comunicação entre os participantes. Alguns deles possuem especialmente muitas interações: um membro obteve respostas de 37 pessoas diferentes. Outro respondeu 39 pessoas. No entanto, a grande maioria dos indivíduos interagiu de forma bastante branda, repondo pouquíssimas pessoas e sendo respondido por pouquíssimas (34 pessoas escreveram 2 respostas ou menos). Este tipo de distribuição é comum em fenômenos sociais<sup>3</sup>.

A comunidade presente na lista de emails da Rede Metareciclagem possui um caráter de atuação específico em parte abarcado pelas ideias de apropriação tecnológica, autonomia, descentralização e *metareciclagem* propriamente dita (que é o ciclo de produção com documentação do processo, para que seja ‘metareciclado’). Essa atuação se dá em grande parte fora dos emails e não pode ser convenientemente recuperada através de repositórios para controle de versão, como no caso da produção de código computacional. Pode-se buscar outros vestígios em blogs e redes como *LinkedIn*. No entanto, esse trabalho se ocupa somente com as estruturas topológicas da rede dos emails. Valioso atentar para a altíssima correlação encon-

trada entre as atividades na lista e as atividades fora da lista, nos casos em que se buscou este cruzamento. Por exemplo, os mais ativos nas listas do *Apache* e do *gcc* são de fato os mais ativos nos repositórios de códigos de cada um dos projetos<sup>2</sup>. Isso pode ser um indício de que a instância aberta e pública se desenrola de forma a dar respaldo e lastro para os mais presentes nas atividades tidas como cruciais pela comunidade. Ao mesmo tempo, dada a continuada e intensa investida dos membros mais ativos, este empenho só incide em quem está realmente envolvido com a proposta, i.e., a participação está calcada em empenho real, até pela repercussão, no próprio indivíduo, das mensagens trocadas.

O objetivo desta pesquisa é estudar as atividades de comunicação e coordenação social humana, usando como dados a troca de mensagens de email entre participantes de uma lista aberta (pública). Como exemplo de aplicação, um participante interessado pode com isso comprovar sua dedicação e conhecimento sobre a comunidade e seus propósitos ou propor atividades dada a melhor compreensão das propriedades globais da rede em que se insere. São de interesse as perguntas:

- Quais as propriedades topológicas da rede social formada com os emails considerados?
- Os membros que mandam muitas mensagens na lista são também dedicados nas outras atividades propostas?
- Programadores e desenvolvedores de tecnologias livres realizam um papel diferente na rede social?
- Os membros mais ativos são os mais influentes na rede?

Para lidar com estas questões, faz-se necessária a obtenção dos dados e a correta associação dos atributos das mensagens para a formação da rede em questão.

## III. SOBRE CÃES E METARECICLEIROS: TRATAMENTO DOS DADOS

“Na internet, ninguém sabe se você é um cachorro”, i.e., a verificação de identidade na autoria dos emails é difícil. Um caso importante é o da utilização de diferentes emails pela mesma pessoa, conhecido como o problema do *alias arbitrário*. Para tratar disso, é útil a clusterização pelo email e pelo nome, a partir do qual, nas comunidades obtidas, os aliases são facilmente detectados. A descrição exata deste procedimento está planejada para um momento próximo, pois não se mostrou necessário para o caso dos 1000 emails iniciais da Rede Metareciclagem, e será necessário no momento da utilização de mais mensagens. Foram 71 autores os que escreveram as 1000 mensagens em um total de 1 mês e 22 dias (de 1 Aug 2005 05:29:59 -0200, uma segunda-feira até 23 Sep 2005 01:27:17 +0000, uma sexta-feira).

só software?  
R: não só  
software.  
Eae? Muito  
difícil de  
entender que  
é um caso  
para-  
digmático?

Frase  
estranha. R:  
melhorou?

Os arquivos públicos da lista disponibilizam quase 58 mil mensagens até o presente momento.

Estas mensagens são acessíveis através de chamadas HTTP. Os serviços do GMANE<sup>4</sup> fornecem as mensagens em formato *mbox*. A seguinte URL, se utilizada em um *browser* comum ou para uma chamada HTTP por outra via, retorna as mensagens (observe o caracter '\ ' para quebra de linha):

#### URL para obtenção via requisição HTTP das mensagens numeradas de 1456 até 2200 em formato mbox.

```
1 wget http://download.gmane.org/gmane.politics.\
2 organizations.metareciclagem/1456/2200
```

Embora bastante eficiente, o processo não é preciso. Quando requisitadas várias mensagens, há o risco de não serem todas baixadas por limites de tamanho do pacote ou outras razões ainda desconhecidas. Por exemplo: as mensagens de 23450 a 24947 estão vazias, não estão acessíveis via URL e o *wget* forma um arquivo de tamanho zero. Assim, este pequeno script em Python baixa todas as mensagens, uma a uma, das quase 58 mil mensagens públicas:

#### Pequena rotina para baixar mensagem por mensagem. Necessária para assegurar consistência dos dados.

```
1 import os
2
3 for i in xrange(1,58000):
4     os.system("wget http://download.gmane.org/\
5 gmane.politics.organizations.metareciclagem/\
6 %i/%i -O %i" % (i,i+1,i))
```

Estes arquivos possuem todos os dados usados neste trabalho. Para a construção da rede, é conveniente utilizar o formato *mbox* e carregar as mensagens, mas pode-se compreender os arquivos e extrair as relações diretamente. Eis um exemplo de *script* que abre os arquivos que estiverem no diretório em que ocorre a execução, desde que sejam os arquivos com as mensagens:

#### Carregamento das mensagens em Python

```
1 import mailbox, os
2
3 a=os.listdir(".")
4 a.sort()
5
6 mm=[]
7 for aa in a:
8     mbox = mailbox.mbox(aa)
9     # versao em plain text,
10    # mesmo quando vier multipart
11    m=mbox[0]
12    mm.append(m)
```

Uma mensagem  $m = mm[i]$ ,  $\forall i \in \mathbb{N}$  é também um dicionário em Python. As chaves '*In - Reply - To*', '*References*' e '*Message - ID*' possuem informação suficiente para estabelecer a relação de resposta de uma à outra.

#### Árvore das mensagens

```
1 In [87]: mm[40]['References']
2 Out[87]: '<bf76ddfc0508100618daf4b5a@mail.gmail.com>\
3 \n\t<bf76ddfc05081006455623b71d@mail.gmail.com>'
4
5 In [88]: mm[40]['In-Reply-To']
6 Out[88]: '<bf76ddfc05081006455623b71d-JsoAwUISXosN+\
7 BqQ9rBEUg@public.gmane.org>'
```

A exemplo da passagem em Python acima, a chave '*References*' possui todas as IDs da sequência de respostas até a mensagem que iniciou a thread. A última ID é a da mensagem que impulsionou a mensagem atual como resposta. Utilizar '*In - Reply - To*' é um pouco mais trabalhoso, a relação de resposta é dada pela *string* até o último sinal de menos ('-') tanto da ID que esta chave retorna quanto da ID que se quer comparar. Várias outras informações estão disponíveis, incluindo data e hora do envio, assunto, nome de quem enviou e corpo das mensagens.

Os resultados são consideravelmente robustos e todas as mensagens foram processadas sem erros. Como validação e para aprofundamento, um pequeno *script* cria o grafo no NetworkX. Este exporta GML (*Graph Modeling Language*) que abre no *software* de visualização de redes Gephi<sup>5</sup>. Os resultados de grau e força de saída e entrada foram consistentes na comparação entre a versão em NetworkX/GML/Gephi e a extração destes dados na força bruta, garantindo a boa formação da rede e os resultados a seguir.

## IV. ANÁLISE DOS RESULTADOS OBTIDOS

A figura 1 mostra um histograma de envio de mensagens. A grande maioria das pessoas mandou somente uma mensagem ou pouquíssimas (44 pessoas mandaram menos de 8 mensagens cada). Destas, quase todas mandaram 1-3 mensagens. Há, ao mesmo tempo, pouquíssimos participantes que enviaram muitas mensagens (3 pessoas mandaram mais de um terço das mensagens). Um número moderado de participantes conta os intermediários (24 usuários mandaram  $\approx 30$  mensagens cada). A figura 2 esboça um histograma do número de respostas obtidas. Ambas as contagens apresentam sinais de ser livre de escala, o que é usual para fenômenos sociais<sup>6</sup>. Além disso, esta é uma propriedade valiosa para um aprofundamento do ponto de vista das ciências naturais. Uma confirmação desta característica pode provir do uso de mais mensagens e de uma apreciação mais cuidadosa das medidas expostas no Apêndice.

Os graus de entrada e saída são mais relacionados a medidas de redes sociais propriamente ditas. O grau de saída é o número de pessoas diferentes que responderam ao participante, um sinal de influência pois suas mensagens repercutiram e um número de pessoas se atentaram ao ponto de escrever algo que lhes ocorre. O grau de entrada é o número de pessoas diferentes a quem o participante respondeu. É um indicador de engajamento

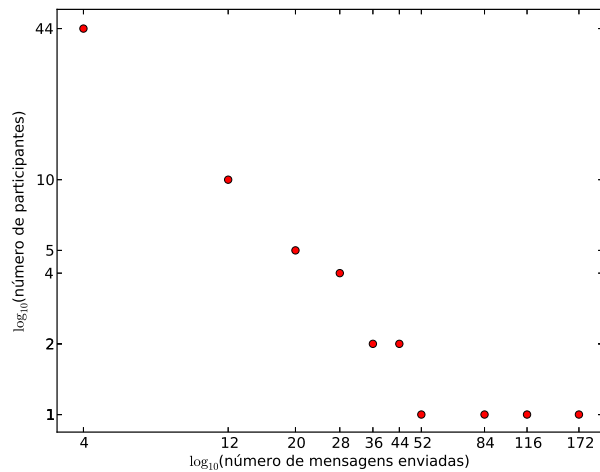


Figura 1. Número de mensagens x número de participantes que enviou a quantidade. O agrupamento foi feito no intervalo de 8 em 8 mensagens. A propriedade livre de escala está sugerida e pode ser confirmada com a utilização de mais dados: pouquíssimos participantes mandaram muitíssimas mensagens, muitíssimos participantes mandaram pouquíssimas mensagens, alguns participantes mandaram algumas mensagens.

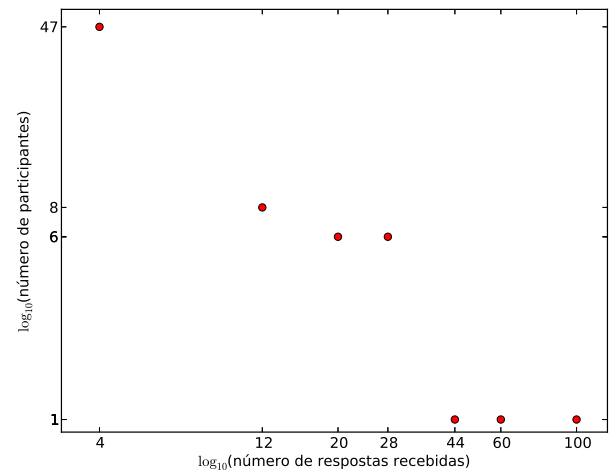


Figura 2. Número de participantes x número de respostas obtidas por cada um. Agrupamento de 10 em 10 mensagens. Há traços da propriedade livre de escala, cuja presença pode ser confirmada com o uso de mais mensagens.

nos propósitos da lista e da força dos interesses do participante. Estes também apresentaram sinais de uma distribuição livre de escala. Omite-se aqui a visualização por encurtamento, dada a ainda pequena quantidade de dados utilizada para que o procedimento tenha validade estatística. Ademais, o âmbito de variação de grau não ultrapassa 40 unidades. Veja o Apêndice para as forças de entrada, saída e somadas assim como os graus de entrada, saída e somados para cada nó da rede.

A observação de simetrias nas relações de resposta merece aprofundamento e talvez seja inédita. Em primeiro lugar, as conexões são altamente assimétricas, i.e. caso haja uma aresta com peso alto de  $a$  para  $b$ , a probabilidade é maior que 0.5 de que não haja uma aresta forte de  $b$  para  $a$ . A figura 3 permite a observação desta propriedade em particular: 10 de suas 14 conexões são assimétricas ao ponto de serem eliminadas do diagrama por um limiar de corte. Estas assimetrias sugerem circuitos mais longos de propagação de informação, envolvendo mais participantes e amenizando os processos herméticos nas comunidades.

A figura 4 é um grafo completo formado pelas 1000 mensagens. Foram removidos os 6 indivíduos que não receberam respostas e não responderam ninguém, pois não formaram aresta com ninguém. A coloração explícita as divisões de níveis de atividade, e, com isso, um aspecto das funções dos indivíduos da rede.

Um aprofundamento analítico das medidas está previsto e estas figuras sugerem propriedades passíveis de maior averiguação. Os dados estão disponibilizados no

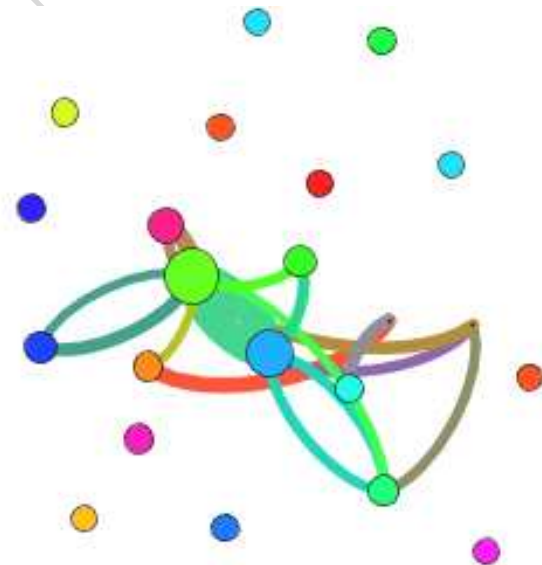


Figura 3. Rede das primeiras 1000 mensagens da Rede Metareciclagem (Ago-Set/2005), com vértices coloridos segundo detecção de modularidade[TODO: definir modularidade]. Arestas com peso baixo não foram mostradas. A maior parte das relações – aproximadamente 71.5% – são assimétricas. Ou seja, em mais de dois terços das relações, uma pessoa responde à outra em intensidade bastante diferente. Isso favorece dinâmicas de espalhamento de informação, aumentando o número de indivíduos dos circuitos e amenizando relações mais fechadas.[Vértices e arestas representam o quê? TODO: figura em formato vetorial para melhor qualidade.]

nao definiu  
forças ainda!  
R: vale  
escrever  
ASAP.

Não foi  
falado sobre  
grafos ainda!  
TODO:  
definir!



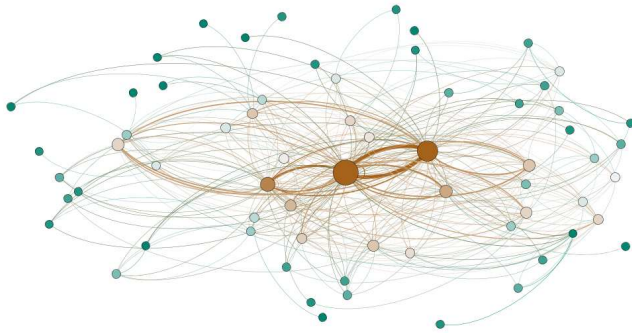


Figura 4. Rede das primeiras 1000 mensagens públicas da Rede Metareciclagem, vértices segundo a centralidade de força que apresentam: os de peso alto se apresentam centralmente e são chamados *hubs*. Os vértices de cor clara são intermediários, tidos como os mais influentes na propagação de opiniões e informação pela literatura. Os nós mais periféricos estão em verde, fundamentais para manutenção da diversidade e para resistência da rede em caso de alguma epidemia.

Apêndice.

## V. CONCLUSÕES E PREVISÕES

A realização das redes de comunicação social humana pelo uso de dados públicos de *email*, a verificação de traços livres de escala e as novas observações afirmam um primeiro passo da análise envisada; apontam para comportamentos naturais já reconhecidos por sua complexidade e do qual a presença de comunidades é marcante. A verificação das relações assimétricas, assim como a comparação entre grau e força, indicam propriedades valiosas. Como exemplos em especial, a relação  $\frac{\text{grau de entrada}}{\text{força total}}$  pode ser compreendida como um índice de promiscuidade e a relação  $\frac{\min[\text{força de } (a,b), \text{força de } (b,a)]}{\max[\text{força de } (a,b), \text{força de } (b,a)]}$  pode ser utilizada como índice de simetria de um par conectado de vértices. Tomando a quantidade  $|\text{força de } (a,b) - \text{força de } (b,a)|$  como a assimetria da relação entre  $a$  e  $b$ , a assimetria do vértice é a soma de todas as assimetrias das relações que mantém. Até onde averiguado, estas relações ainda não foram estudadas.

O uso reduzido de mensagens permitiu um início consistente, mas diminuiu a assertividade sobre as proprie-

dades livres de escala e de pequeno mundo. O uso de um conjunto maior de mensagens poderá confirmar as propriedades, e revelar sutilezas de como elas se manifestam. O cruzamento das diferentes listas por estes métodos se mostra pertinente para a compreensão de aspectos gerais e específicos das incidências.

São vislumbradas aplicações para fluxo de bens e de informação, assim como para dinâmicas deliberativas, o que pode vir a nortear nossos esforços. A criação de jogos que usufruam destes conhecimentos pode facilitar uma apropriação lúdica e o desenvolvimento de tecnologias. Pode-se também utilizar técnicas de aprendizado de máquina para agrupamento e detecção de comunidades e contextos similares em redes sociais.

Informações acerca das *threads* iniciadas, como número de *threads* iniciadas por cada participante, número de respostas que cada *thread* recebeu, espaçamento de tempo entre as mensagens, localizações geográficas, gêneros e os conteúdos verbais dos assuntos e corpos das mensagens: todas estas características estão disponíveis publicamente para consulta e estudos. O cruzamento destas informações com dados provenientes de fora das mensagens de emails é particularmente atraente no caso dos repositórios e do IRC<sup>7</sup>, permitindo caracterizar uma lista como de desenvolvedores e estabelecer outros vínculos.

Tratamentos da própria rede podem facilitar a extração de características. Por exemplo: os hubs estão fortemente conectados entre si, então pode-se unificá-los em um só vértice. Outro exemplo pertinente é mapear a rede para uma rede acíclica, o que explicita propriedades visual e analiticamente. Foram verificadas variações das propriedades com o passar do tempo tanto para a rede como um todo quanto para vértices que desempenham papéis especiais, o que merece um estudo em particular.

O trabalho feito em diálogo com a lista de emails Metareciclagem potencializa esta investida com novas ideias e amadurecimento. O contato com a comunidade possibilita uma série de análises voltadas para o usuário sobre os próprios usuários e sobre a rede como um todo.

## AGRADECIMENTOS

Vilson Vieira e Renato Fabbri são gratos à CAPES e ao comitê de pós-graduação do IFSC. Renato Fabbri é grato ao Vilson Vieira, Ricardo Fabbri, Glorm Soares, Silvio Shefa, AVAV, Cláudia Fornari, Edson "Presto" Correa, Robson Cayley, Ricardo Poppi e Patrícia Ferraz pelo apoio a este trabalho. Os autores agradecem à Rede Metareciclagem e aos membros da lista de emails da rede. Agradecem também aos Profs. Luciano da Fontoura Costa e Osvaldo Novais de Oliveira Junior, ambos do IFSC/USP, pelos trabalhos de referência que realizaram e são aqui utilizados.

não foram  
definidos  
hubs neste  
artigo

## REFERÊNCIAS

- <sup>1</sup>K. Marek-Spartz, P. Chesley, and H. Sande, "Construction of the gmane corpus for examining the diffusion of lexical innovations," (2012).
- <sup>2</sup>C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan, "Mining email social networks," in *Proceedings of the 2006 international workshop on Mining software repositories* (ACM, 2006) pp. 137–143.
- <sup>3</sup>Luciano da F.Costa, O.N. Oliveira Jr, G. Travieso, F.A. Rodrigues, P.R.V. Boas, L. Antiqueira, M.P. Viana, and L.E.C. Rocha, "Analyzing and modeling real-world phenomena with complex networks: a survey of applications," *Advances in Physics* **60** (2011).
- <sup>4</sup>"Gmane is a mailing list archive.." <http://gmane.org/>.
- <sup>5</sup>M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," (2009).
- <sup>6</sup>L. da Fontoura Costa, O.N. Oliveira Jr, G. Travieso, F.A. Rodrigues, P.R.V. Boas, L. Antiqueira, M.P. Viana, and L.E.C. Rocha, "Analyzing and modeling real-world phenomena with complex networks: a survey of applications," *Advances in Physics* **60**, 329–412 (2011).
- <sup>7</sup>C. Simpson, "Internet relay chat.." *Educational Media and Technology Yearbook* **25**, 62–65 (2000).
- <sup>8</sup>l. f. amancio d. r. ; fabbri r. ; oliveira jr. o. n. ; nunes m.g.v. ; costa, "opinion discrimination using complex network features," 2nd international workshop on complex networks(2011).
- <sup>9</sup>l. f. vieira v. ; fabbri, r. ; travieso gonzalo; oliveira jr. o. n. ; costa, "a quantitative approach to evolution of music and philosophy," *journal of statistical mechanics*(2012).
- <sup>10</sup>l. f amancio d. r. ; fabbri r. ; nunes m. g. v. ; oliveira jr. o. n. ; costa, "distinguishing between positive and negative opinions with complex network features," *textgraphs-5 - 2010 workshop on graph-based methods for natural language processing - acl, uppsala*(2010).
- <sup>11</sup>G. Deleuze, *Difference and Repetition* (Continuum, 1968).
- <sup>12</sup>F. de Saussure, *Course in General Linguistics* (Books LLC, 1916).
- <sup>13</sup>A. Papoulis S. U. Pillai, *Probability, Random Variables and Stochastic Processes* (McGraw Hill Higher Education, 2002).
- <sup>14</sup>R. A. Johnson D. W. Wichern, *Applied Multivariate Statistical Analysis* (Prentice Hall, 2007).
- <sup>15</sup>C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Prentice Hall, 1992).
- <sup>16</sup>R. O. Duda P. E. Hart D. G. Stork, *Pattern Classification* (Wiley-Interscience, 2000).
- <sup>17</sup>L. da F. Costa R. M. C. Jr., *Shape Analysis and Classification: Theory and Practice (Image Processing Series)* (CRC Press, 2000).
- <sup>18</sup>Luciano da Fontoura Costa Renato Fabbri, Osvaldo N. Oliveira Jr., "Philosometrics," (October 2010), arXiv:1010.1880v2.
- <sup>19</sup>Joe H Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association* **58**, 236–244 (1963).
- <sup>20</sup>Hal Varian, "Bootstrap tutorial," *The Mathematica Journal* **9** (2005).
- <sup>21</sup>Christian P. Robert, "Simulation in statistics," in *Proceedings of the 2011 Winter Simulation Conference* (2011) arXiv:1105.4823.
- <sup>22</sup>o. n. fabbri r. ; amancio d. r. ; da f. costa luciano ; oliveira junior, "Use of complex networks for the analysis of speech," *Encontro Nacional de Física da Matéria Condensada*(2010).
- <sup>23</sup>Osvaldo Novais de. Fabbri R. Costa L da F. Oliveira Junior, "Redes complexas no processamento de fala," *Workshop da Pós-Graduação do IFSC. Livro de Resumos*(2010).

**Apêndice A: Mensagens enviadas, forças e graus de entrada e saída, por participante**

- O número de mensagens enviadas por cada indivíduo, em ordem decrescente é: 169, 112, 81, 48, 45, 43, 37, 32, 31, 28, 28, 26, 23, 23, 22, 17, 17, 15, 11, 10, 10, 10, 9, 9, 9, 9, 8, 7, 7, 7, 6, 6, 5, 5, 5, 5, 4, 4, 4, 4, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1.

- As forças de saída, i.e. número de respostas obtidas por cada participante, são: 131, 97, 62, 40, 31, 30, 30, 28, 27, 25, 22, 19, 17, 17, 16, 16, 14, 11, 10, 10, 9, 9, 9, 8, 7, 7, 7, 6, 6, 6, 5, 5, 5, 4, 4, 4, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0.

Os participantes, se considerados 0-70 os elencados através da ordem decrescente de mensagens enviadas, é 0, 1, 2, 5, 14, 9, 4, 7, 6, 3, 13, 8, 16, 17, 12, 11, 15, 18, 19, 22, 25, 23, 20, 21, 26, 35, 24, 29, 38, 27, 10, 39, 61, 30, 28, 36, 41, 32, 34, 59, 47, 44, 31, 51, 49, 52, 40, 37, 50, 57, 69, 54, 46, 42, 60, 66, 45, 33, 68, 65, 43, 64, 63, 62, 48, 53, 56, 67, 55, 58, 70.

- As forças de entrada, i.e. número de respostas escritas por cada participante, são: 151, 109, 56, 39, 37, 35, 29, 28, 26, 20, 19, 18, 17, 16, 15, 15, 14, 10, 10, 9, 9, 9, 9, 7, 7, 7, 6, 6, 6, 5, 5, 5, 5, 4, 4, 3, 3, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.

Os participantes são: 0, 3, 24, 5, 19, 34, 25, 7, 1, 9, 11, 37, 32, 55, 15, 14, 17, 6, 18, 13, 12, 8, 4, 16, 2, 27, 10, 29, 21, 28, 33, 22, 20, 23, 26, 30, 31, 38, 36, 35, 39, 40, 44, 42, 43, 47, 45, 41, 48, 49, 50, 46, 52, 51, 53, 57, 60, 58, 54, 62, 59, 56, 61, 63, 67, 65, 64, 66, 68, 69, 70.

- Os graus de saída, i.e. número de pessoas diferentes que responderam cada um, são: 37, 30, 22, 16, 16, 16, 14, 12, 11, 11, 10, 10, 9, 9, 9, 8, 8, 8, 7, 7, 7, 6, 6, 6, 5, 5, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.

Os respectivos participantes são: 0, 1, 2, 7, 4, 5, 9, 3, 8, 13, 15, 14, 16, 12, 11, 22, 6, 23, 18, 26, 19, 20, 17, 21, 25, 29, 38, 39, 30, 28, 61, 27, 10, 24, 41, 32, 34, 35, 44, 47, 36, 59, 31, 52, 49, 51, 40, 46, 60, 45, 57, 33, 54, 64, 65, 66, 37, 50, 42, 68, 69, 43, 55, 58, 63, 56, 53, 62, 48, 67, 70.

- O grau de entrada, i.e. o número de pessoas diferentes a quem o participante respondeu, é de: 39, 33, 23, 18, 16, 14, 14, 14, 13, 12, 12, 12, 12, 11, 10, 9, 8, 7, 7, 7, 7, 6, 6, 5, 5, 5, 5, 4, 4, 4, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.

Os respectivos participantes são: 0, 1, 2, 4, 9, 11, 3, 5, 14, 12, 7, 6, 8, 10, 13, 15, 21, 17, 19, 20, 23,

22, 24, 26, 30, 25, 27, 18, 28, 32, 31, 36, 43, 29, 35, 33, 44, 50, 41, 38, 52, 45, 47, 53, 54, 42, 68, 60, 58, 57, 61, 51, 63, 40, 34, 39, 64, 16, 59, 69, 67, 37, 66, 65, 48, 62, 46, 56, 49, 55, 70.

## **Apêndice B: Medidas derivadas e adicionais**

As páginas a seguir possuem medidas sobre a rede obtida com as primeiras mil mensagens públicas da Rede Metareciclagem.

RASCUNHO



# Connected Components Report

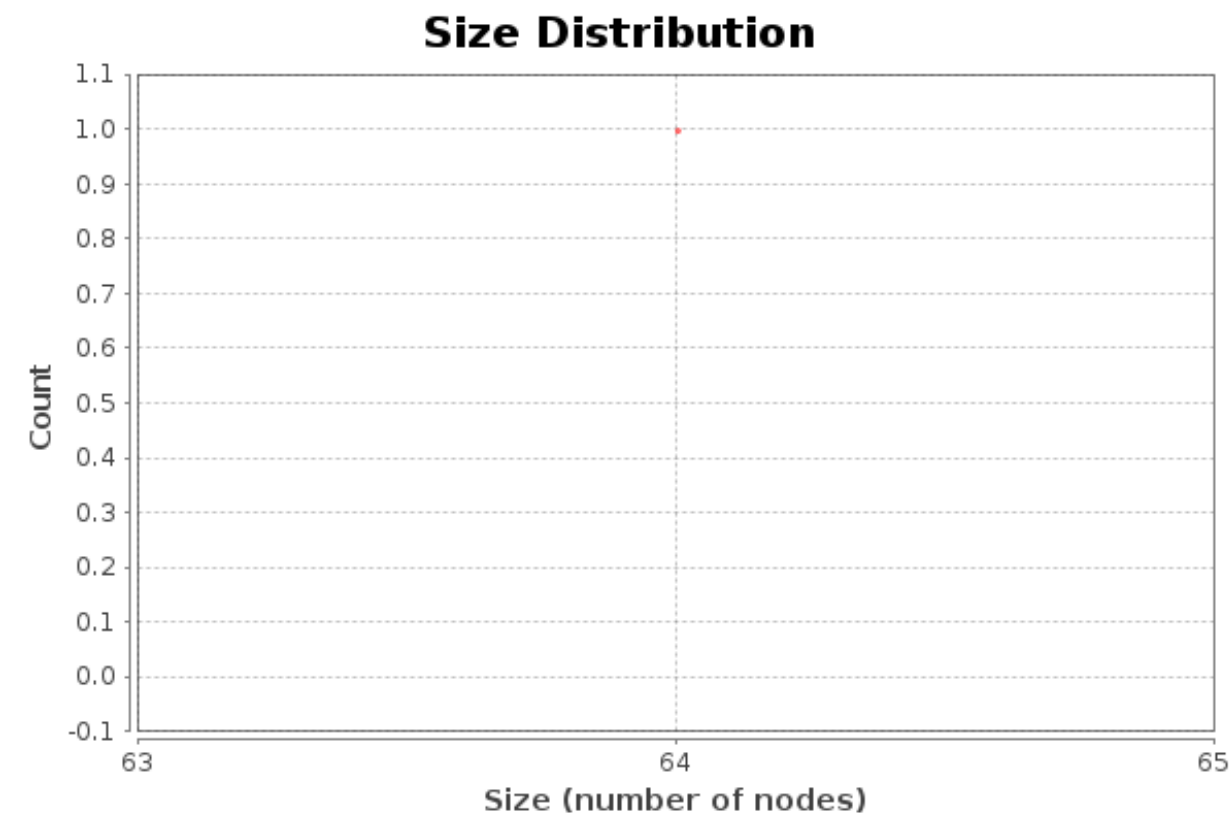
---

## Parameters:

Network Interpretation: directed

## Results:

Number of Weakly Connected Components: 1  
Number of Strongly Connected Components: 14



## Algorithm:

Robert Tarjan, *Depth-First Search and Linear Graph Algorithms*, in SIAM Journal on Computing 1 (2): 146–160 (1972)

# Graph Density Report

---

## Parameters:

Network Interpretation: directed

## Results:

Density:0.096

# Graph Distance Report

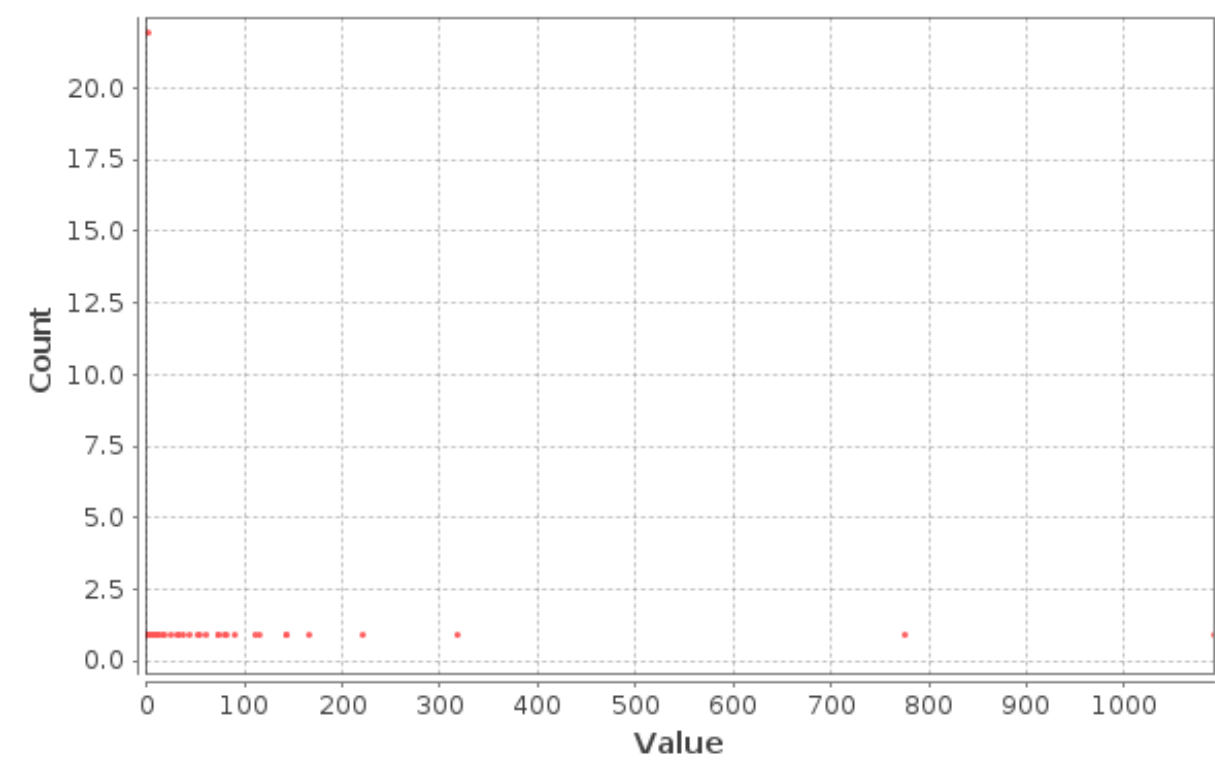
## Parameters:

Network Interpretation: directed

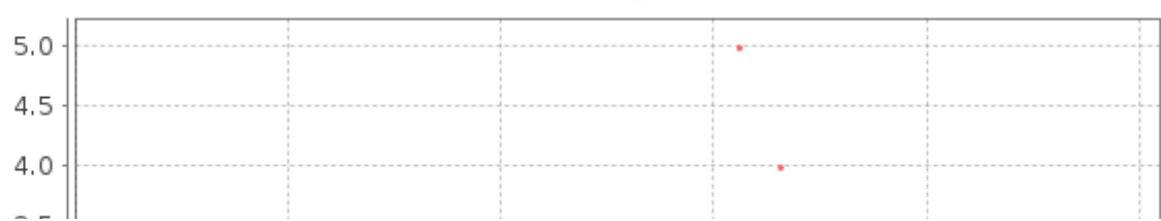
## Results:

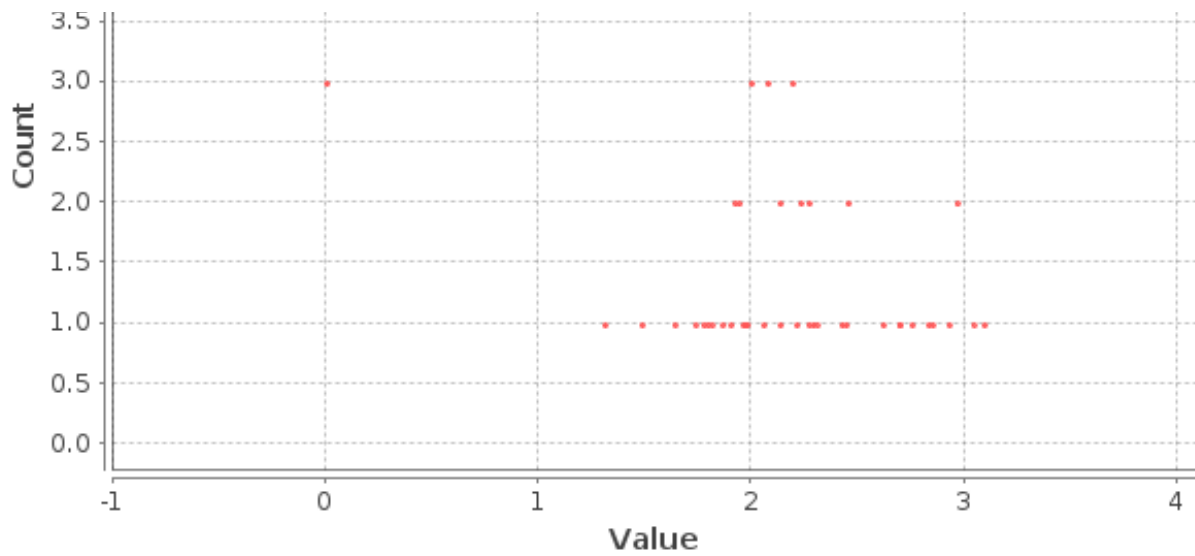
Diameter: 4  
Radius: 0  
Average Path length: 2.217719132893497  
Number of shortest paths: 3183

Betweenness Centrality Distribution

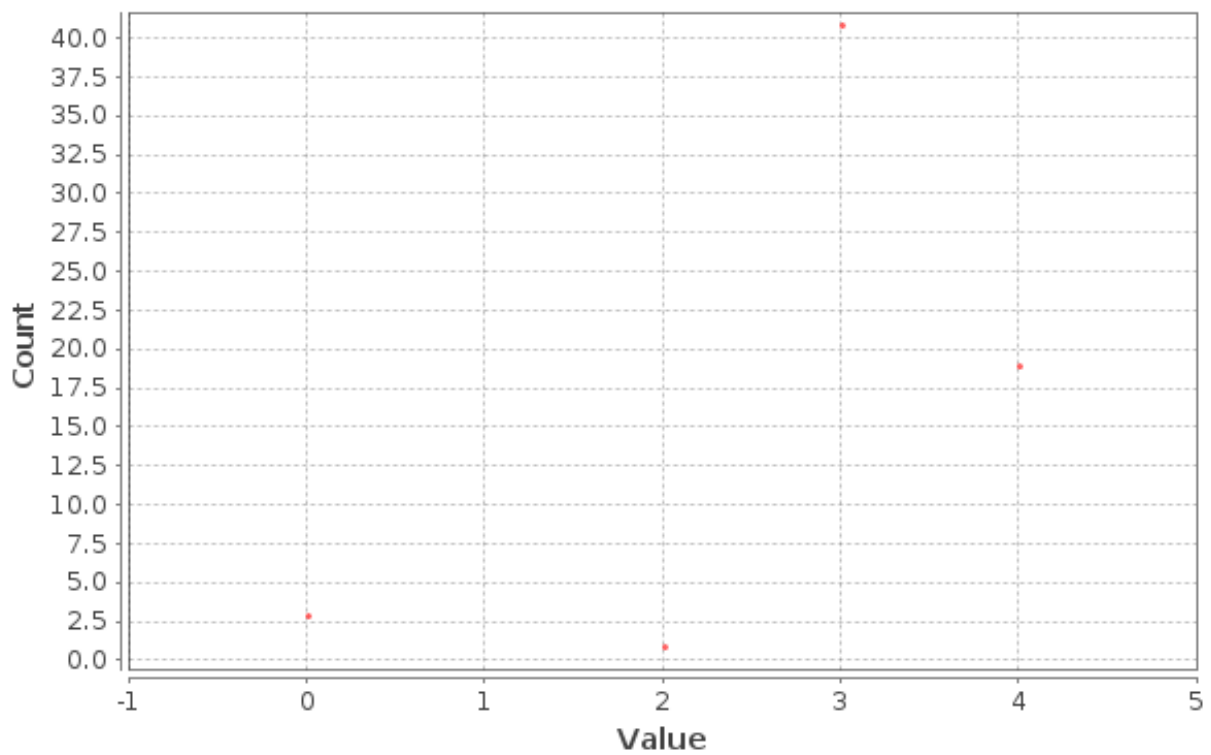


Closeness Centrality Distribution





### Eccentricity Distribution



### Algorithm:

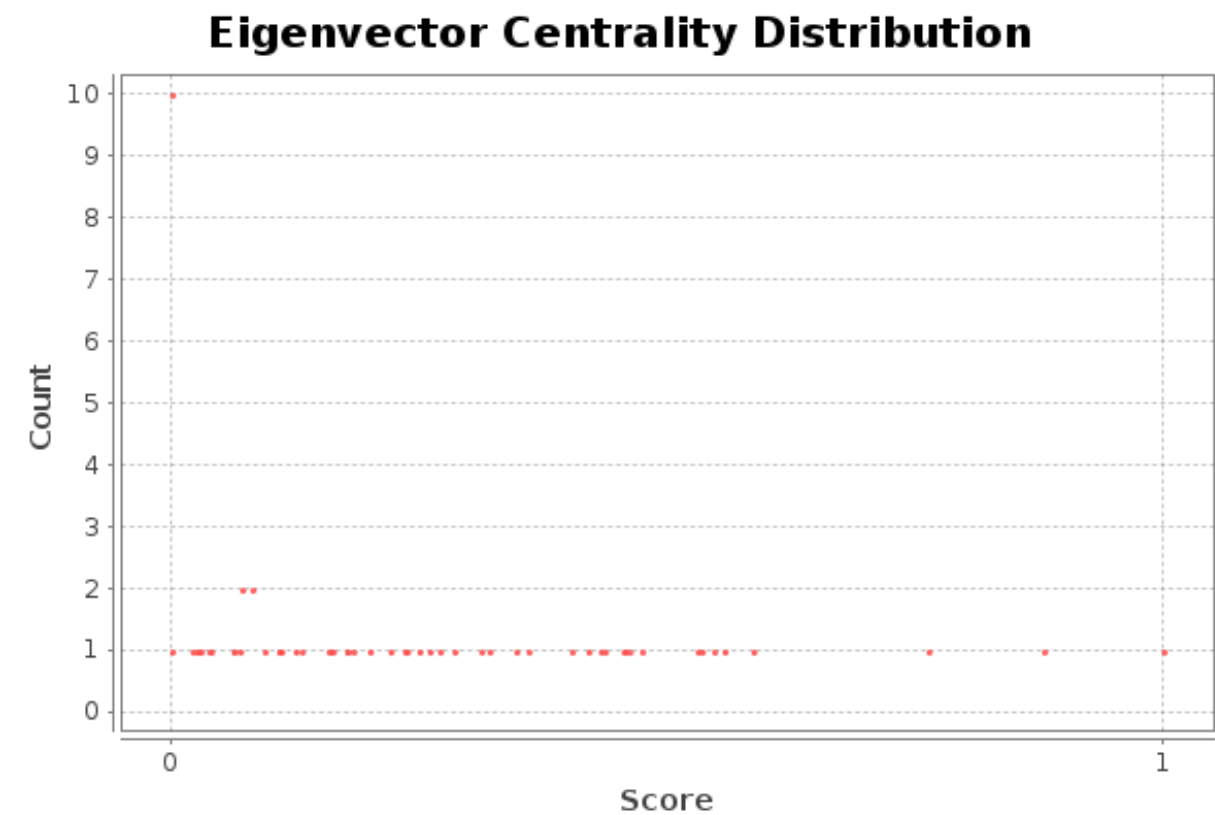
Ulrik Brandes, *A Faster Algorithm for Betweenness Centrality*, in Journal of Mathematical Sociology 25(2):163-177, (2001)

# Eigenvector Centrality Report

## Parameters:

Network Interpretation: directed  
Number of iterations: 100  
Sum change: 8.962086254013716E-4

## Results:



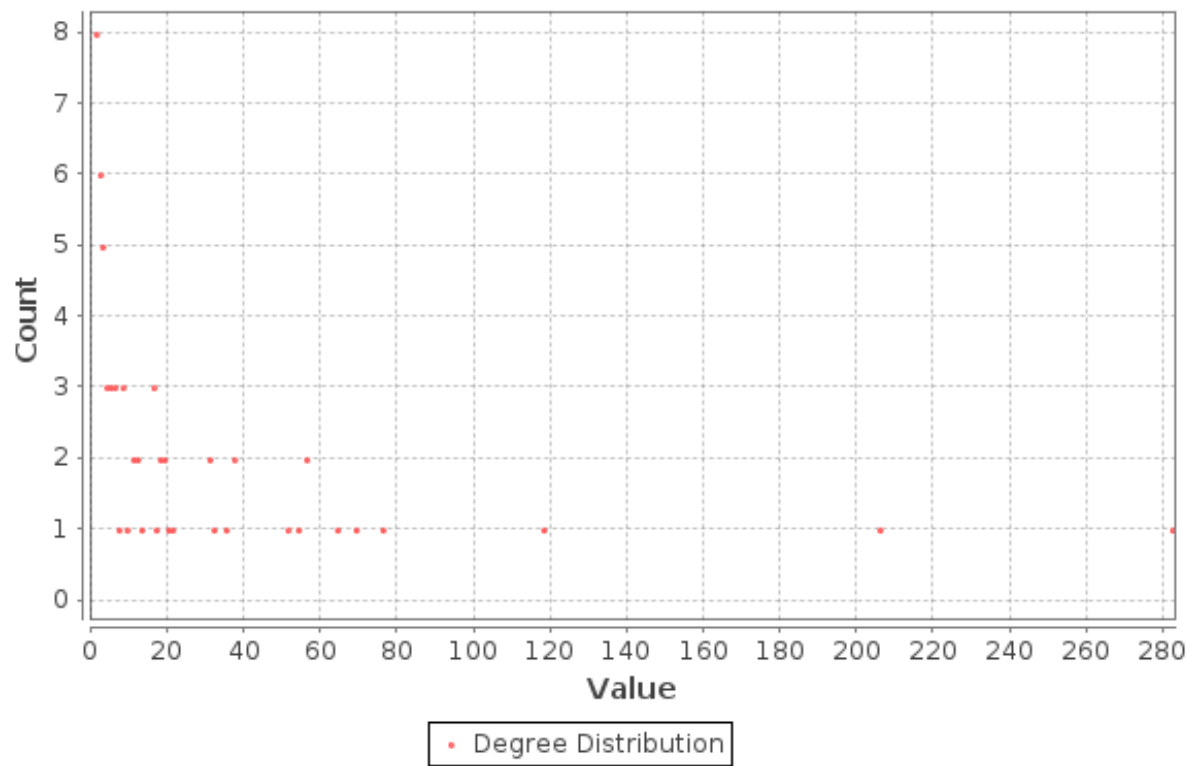


# Weighted Degree Report

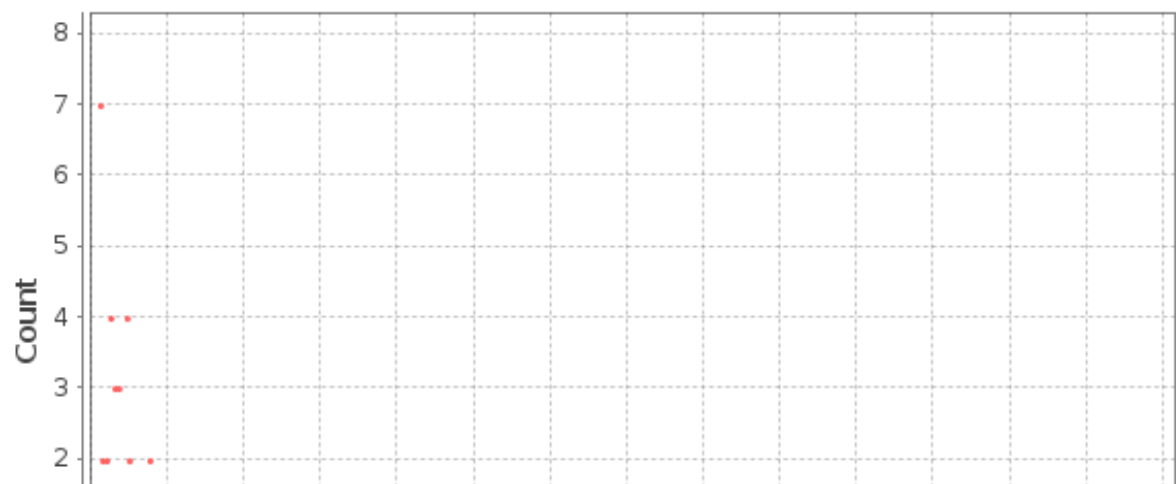
## Results:

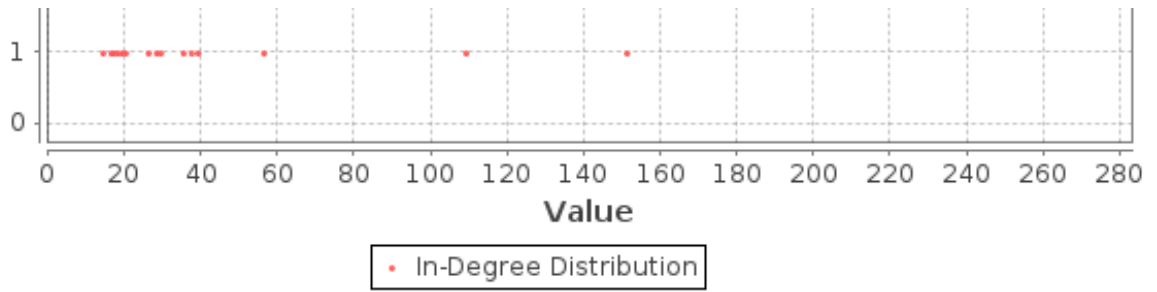
Average Weighted Degree: 12.453

Degree Distribution

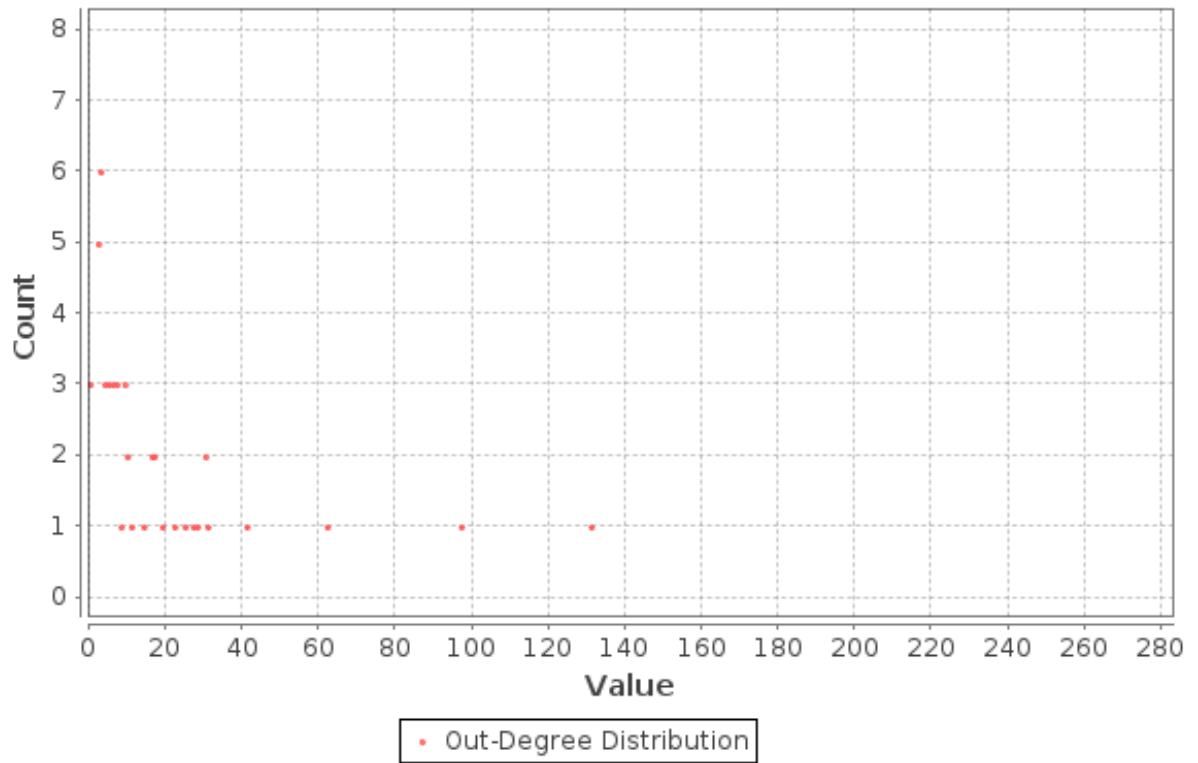


In-Degree Distribution





## Out-Degree Distribution

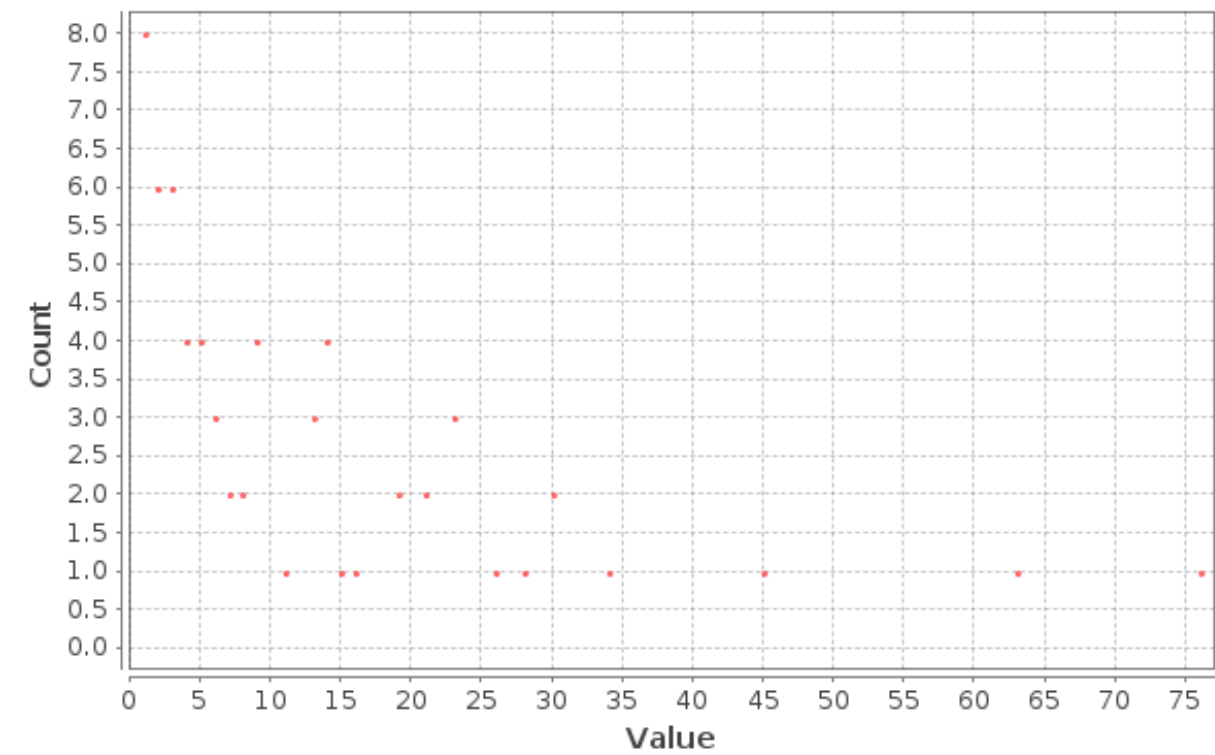


# Degree Report

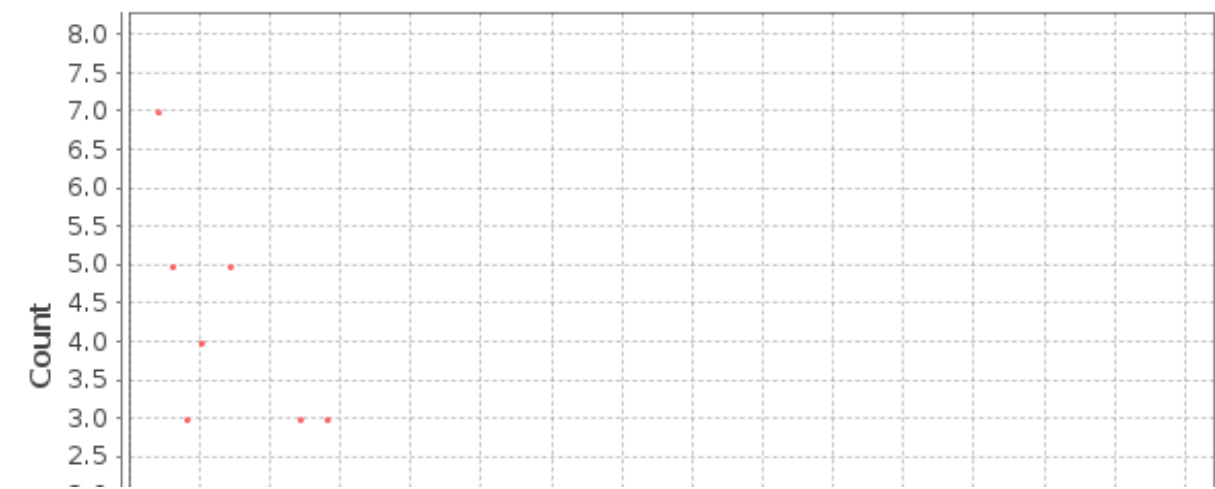
## Results:

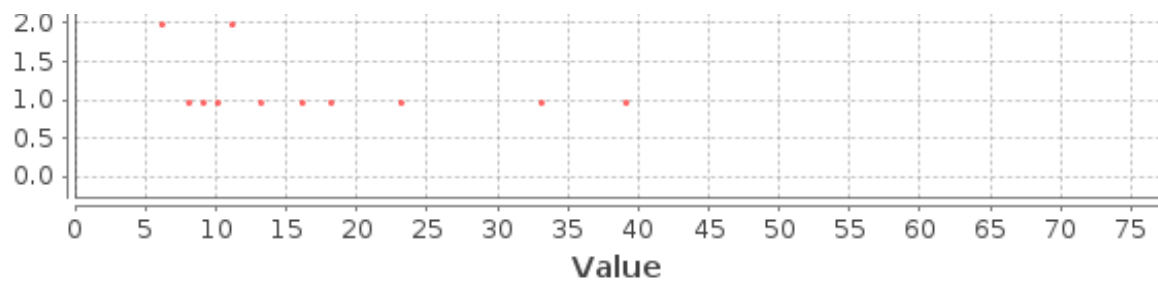
Average Degree: 6.062

Degree Distribution

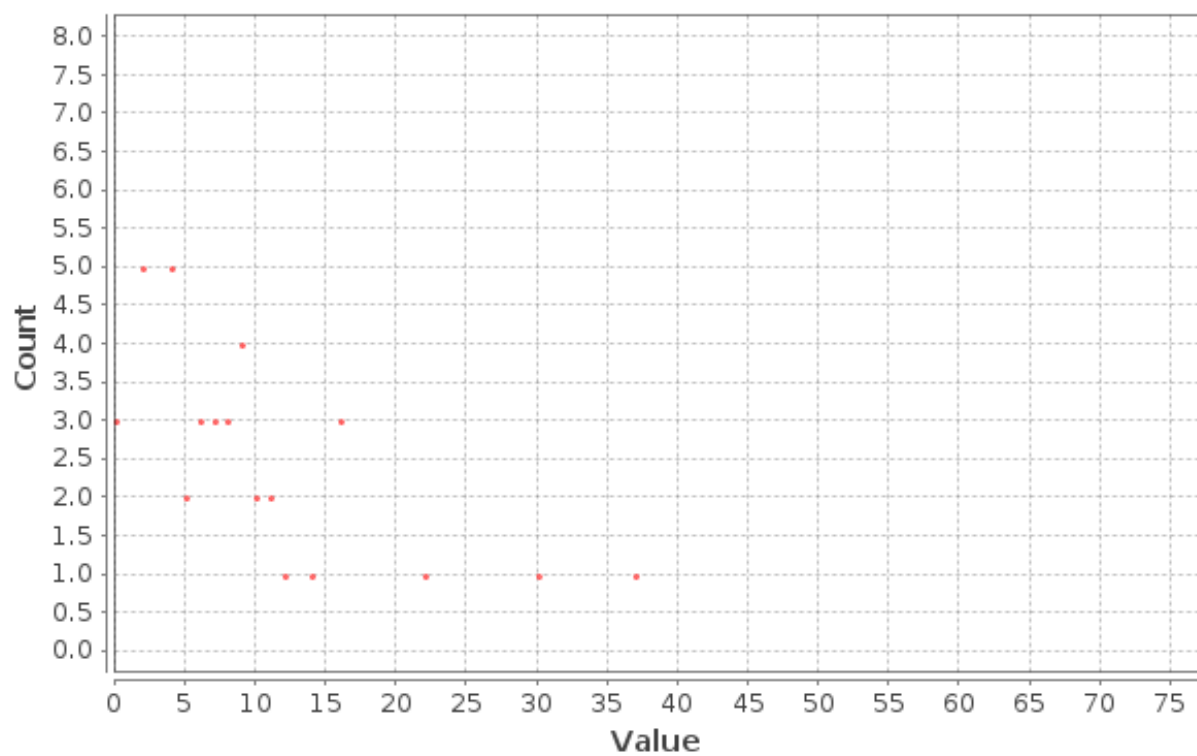


In-Degree Distribution





## Out-Degree Distribution



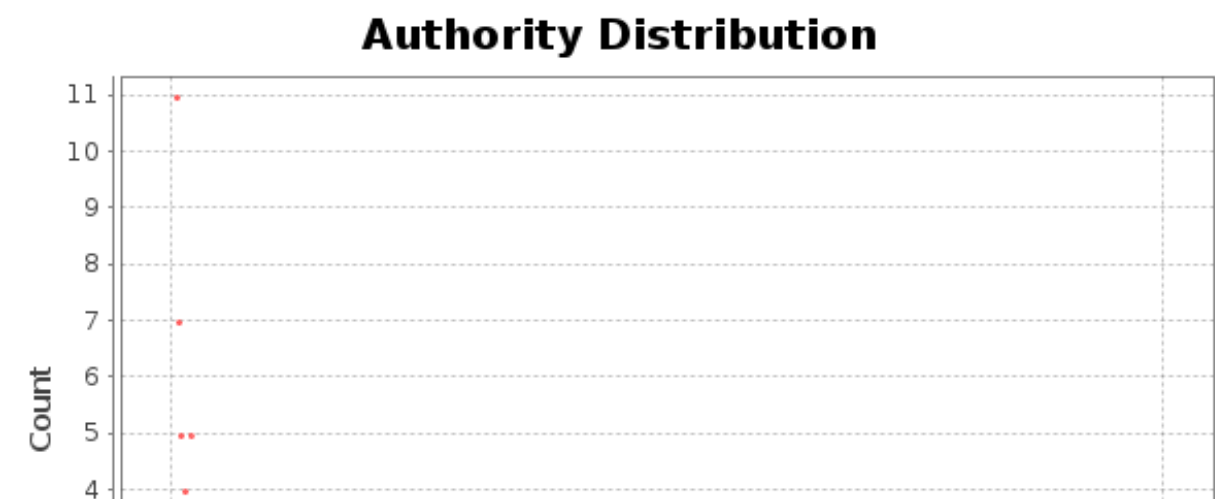
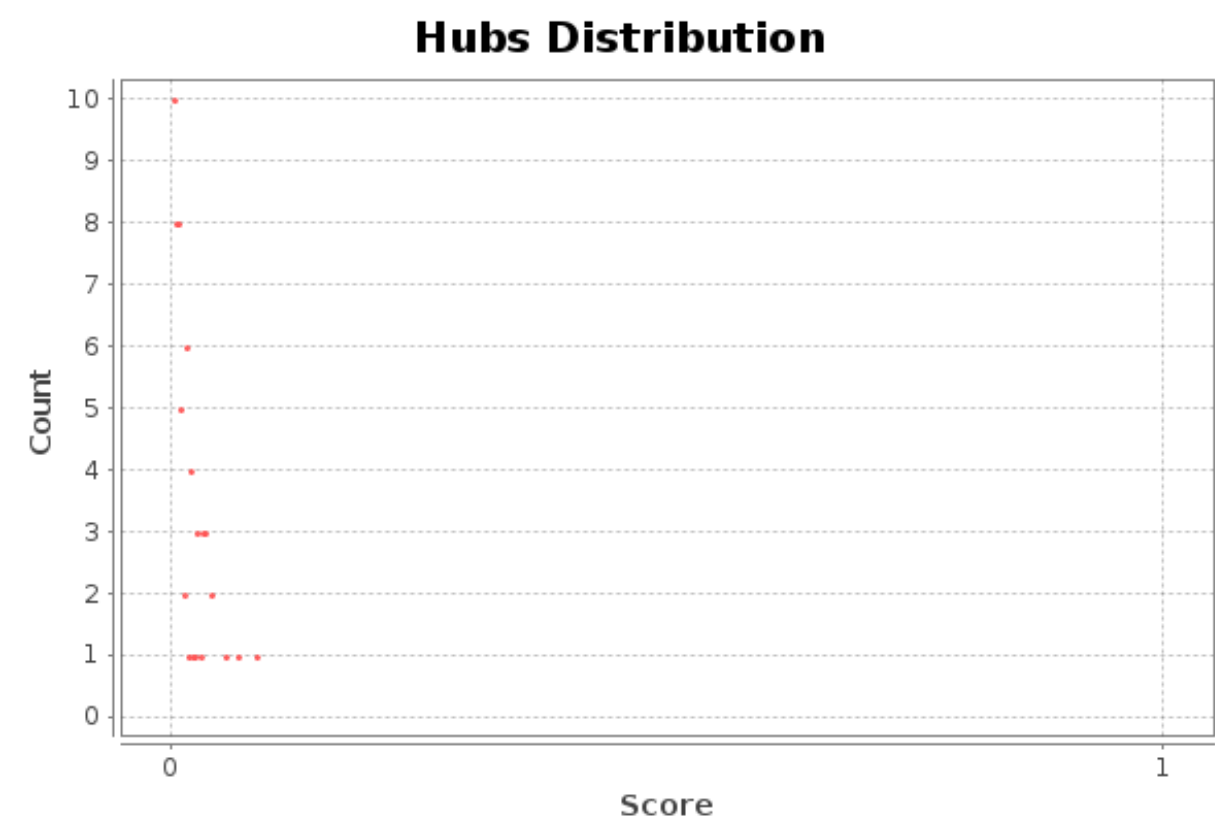
# HITS Metric Report

---

**Parameters:**

E = 1.0E-4

**Results:**







### Algorithm:

Jon M. Kleinberg, *Authoritative Sources in a Hyperlinked Environment*, in Journal of the ACM 46(5):604–632 (1999)

# Modularity Report

---

## Parameters:

Randomize: On  
Use edge weights: On  
Resolution: 1.0

## Results:

Modularity: 0.111  
Number of Communities: 4



## Algorithm:

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, *Fast unfolding of communities in large networks*, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

**Resolution:**

R. Lambiotte, J.-C. Delvenne, M. Barahona *Laplacian Dynamics and Multiscale Modular Structure in Networks* 2009

# Clustering Coefficient Metric Report

---

## Parameters:

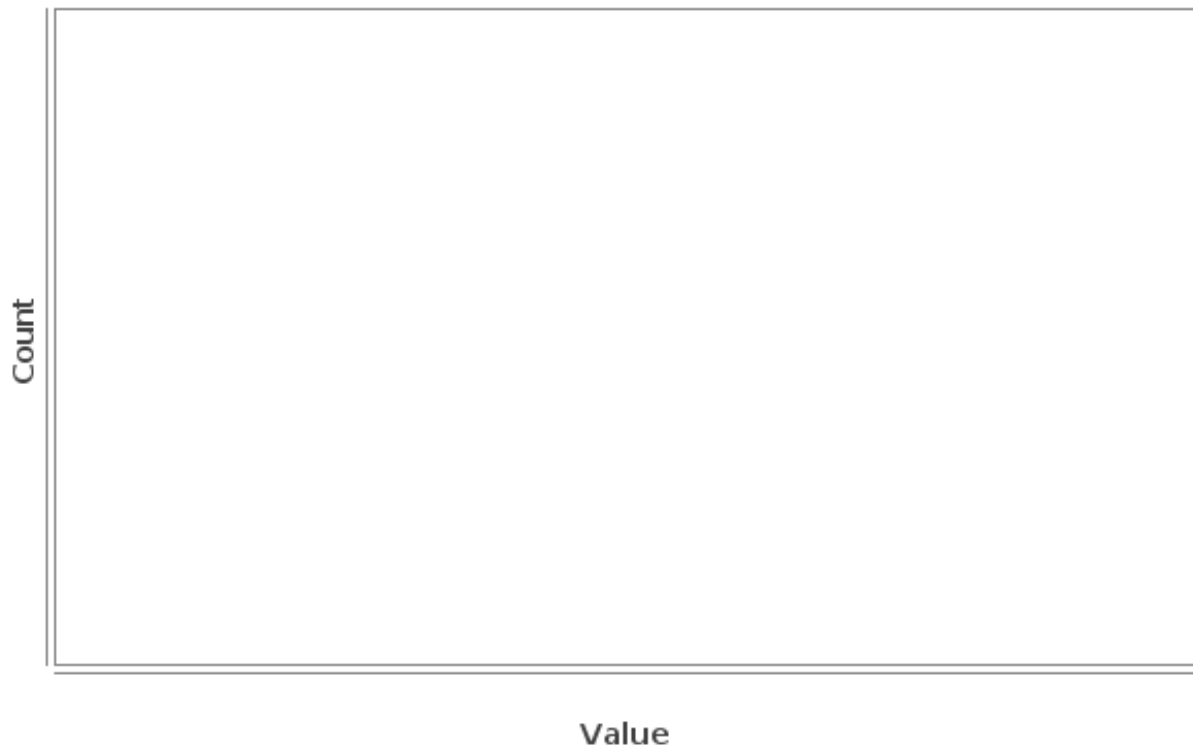
Network Interpretation: directed

## Results:

Average Clustering Coefficient: 0.446

The Average Clustering Coefficient is the mean value of individual coefficients.

## Clustering Coefficient Distribution



## Algorithm:

Simple and slow brute force.

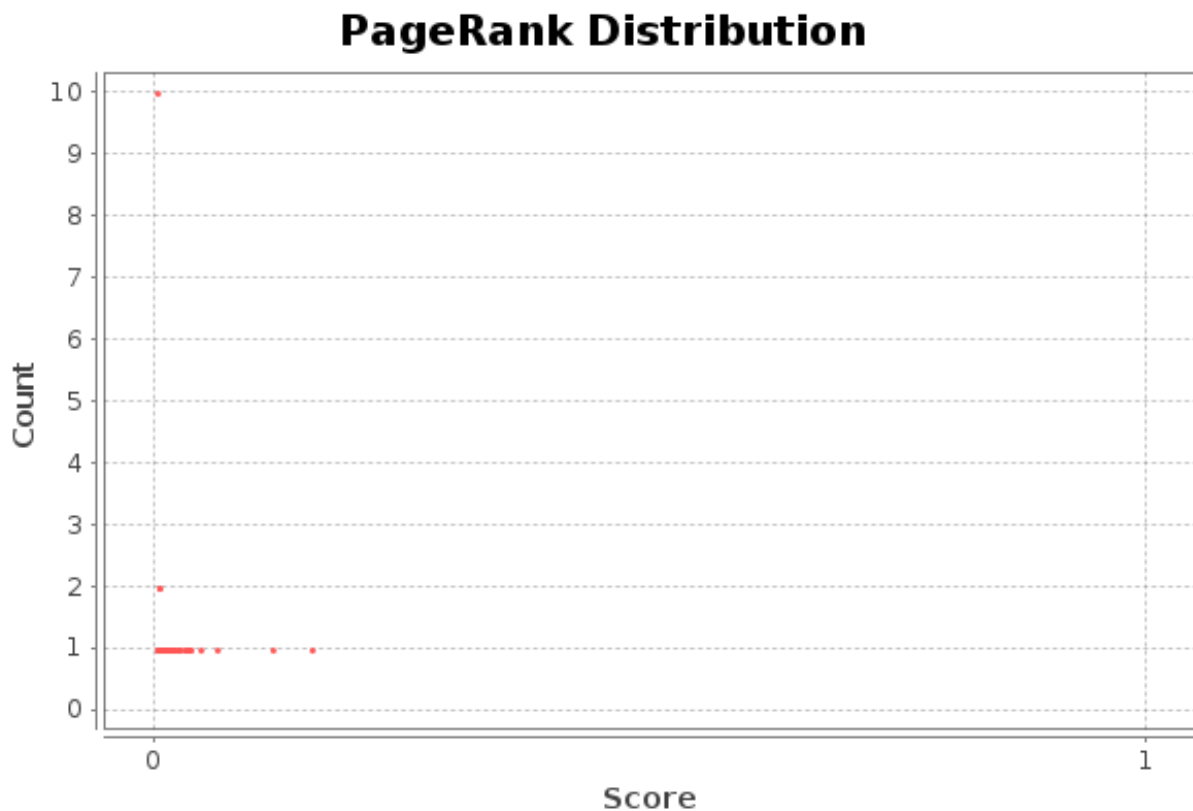
# PageRank Report

---

## Parameters:

Epsilon = 0.001  
Probability = 0.85

## Results:



## Algorithm:

Sergey Brin, Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, in Proceedings of the seventh International Conference on the World Wide Web (WWW1998):107-117