Técnicas de Visualização de Dados aplicadas à Segurança da Informação

André Ricardo Abed Grégio, Benício Pereira de Carvalho Filho, Antônio Montes, **Rafael Santos**

Introdução

Eventos, ferramentas, eventos...

- Lado A:
  - Sistemas computacionais interconectados têm a capacidade de gerar registros das atividades de seus componentes.
  - Registros representam eventos e servem para a monitoração.
  - Mecanismos de defesa geram eventos.

- Lado B:
  - Muitos ataques automatizados “despropositados”.
  - Tráfego prolixo (protocolos de comunicação).
  - Disseminação de código malicioso.
  - Muitas ferramentas gerando *logs* e **não integradas**.

Homens-hora

- Aplicações de correlação de eventos de segurança envolvem trabalho manual.

- Muitas fontes de dados para gerar uma linha coerente de eventos → problemas!
  - Sincronismo;
  - Integridade;
  - Filtragem de eventos não importantes.

- Como “ver” os eventos de forma facilitada?

Visualização!

- Graficamente, pode ser mais fácil identificar padrões em geral.

- Visualização para eventos de segurança está em foco em forense computacional, análise de *malware*, administração de redes, correlação de *logs*.

- Pesquisas geram aplicações para visualizar:
  - Binários
  - Tráfego de rede
  - Chamadas de sistema
  - Registros

Visualização?

- Uma infinidade de maneiras para visualizar eventos de segurança.

- Provêem informações úteis?
- Facilitam a extração de conhecimento?
- Qual a complexidade?
- Melhor que um gráfico de pizza?

- Veremos...

Conceitos de Visualização

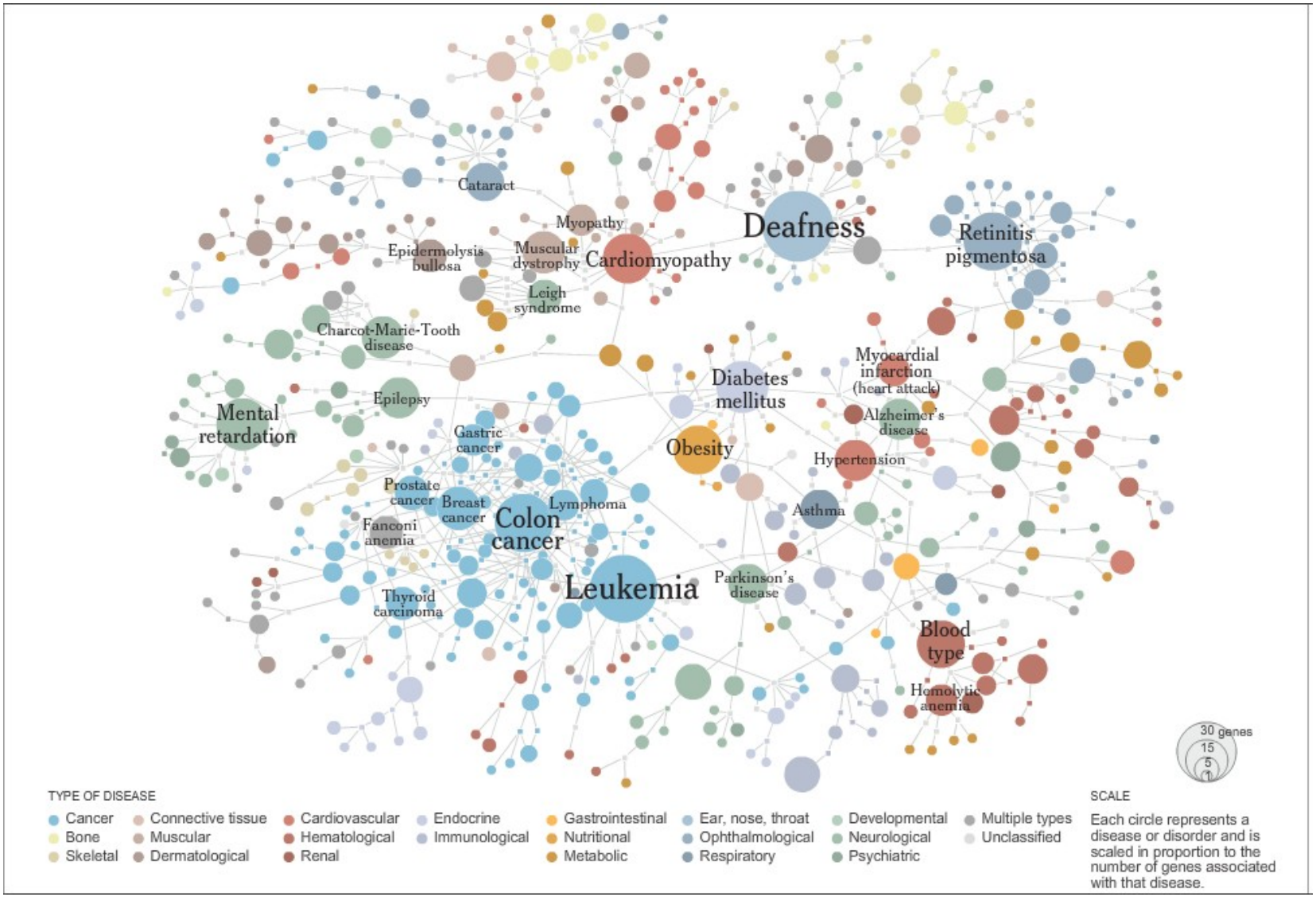
Objetivos de Visualização

- Análise exploratória:
  - Temos os dados, não temos hipótese sobre os mesmos.
  - Busca visual por padrões, estruturas, etc.
- Análise para confirmação:
  - Temos os dados e hipótese sobre os mesmos.
  - Busca visual para confirmar ou rejeitar.
- Apresentação
  - Técnica adequada deve ser usada!

Conceitos de Visualização

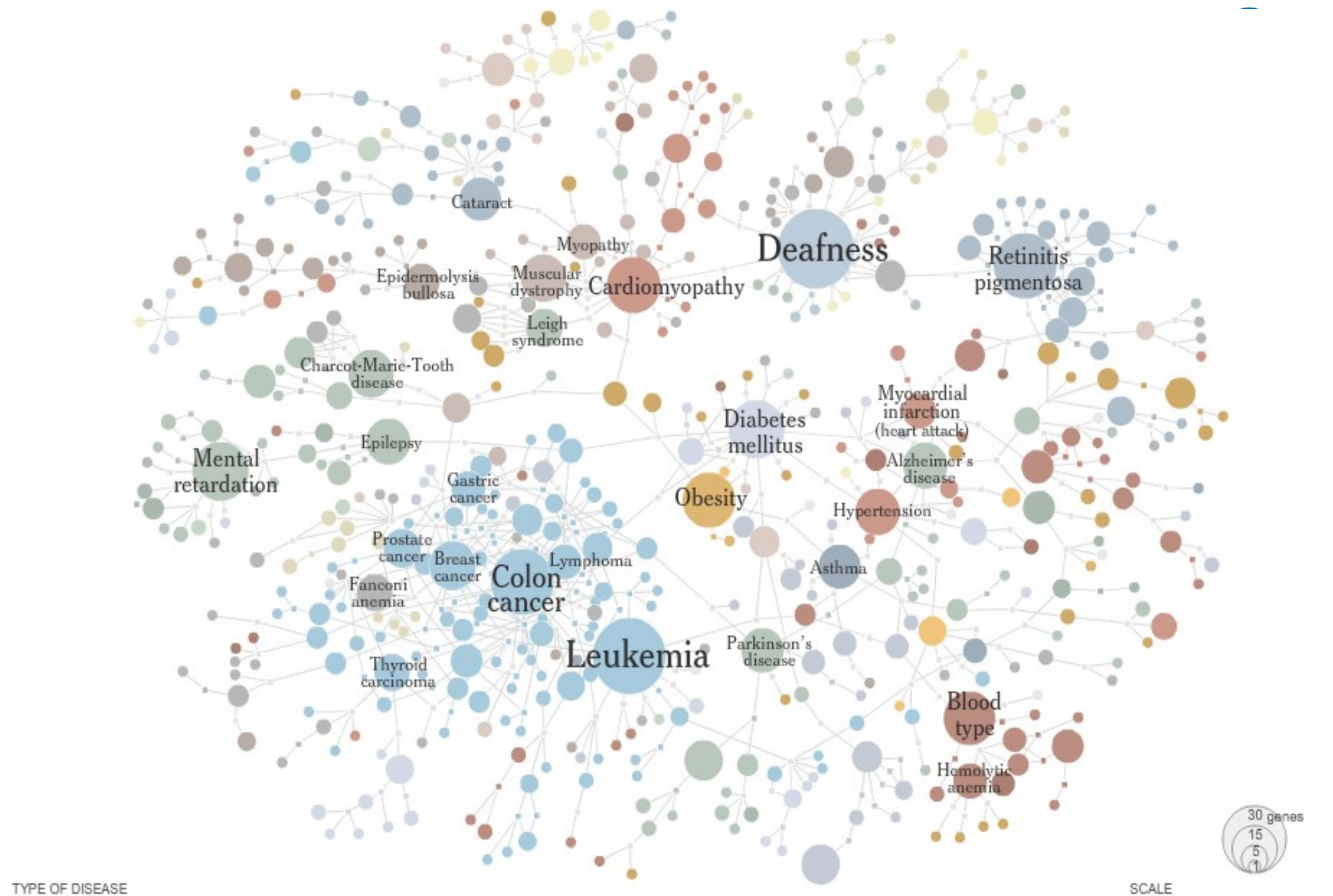
- Edward Tufte, *The Visual Display of Quantitative Information:*
  - "... gráficos sobre dados podem fazer muito mais do que simplesmente ser substitutos para pequenas tabelas estatísticas. Na sua melhor concepção, gráficos são instrumentos para compreender informação quantitativa."
  - "Frequentemente a forma mais efetiva de descrever, explorar e sumarizar um conjunto de números – mesmo um conjunto com muitos números – é **ver figuras destes números**."
  - "Adicionalmente, de todas as formas de analisar e comunicar informação estatística, gráficos bem feitos sobre dados são geralmente ao mesmo tempo a mais simples e mais poderosa".

Supergráficos



http://visualthinkmap.ning.com/

Mapping the Human 'Diseasome' by Marc Vidal, Albert-Laszlo Barabasi and Michael Cusick: ligação entre doenças e genes em comum.



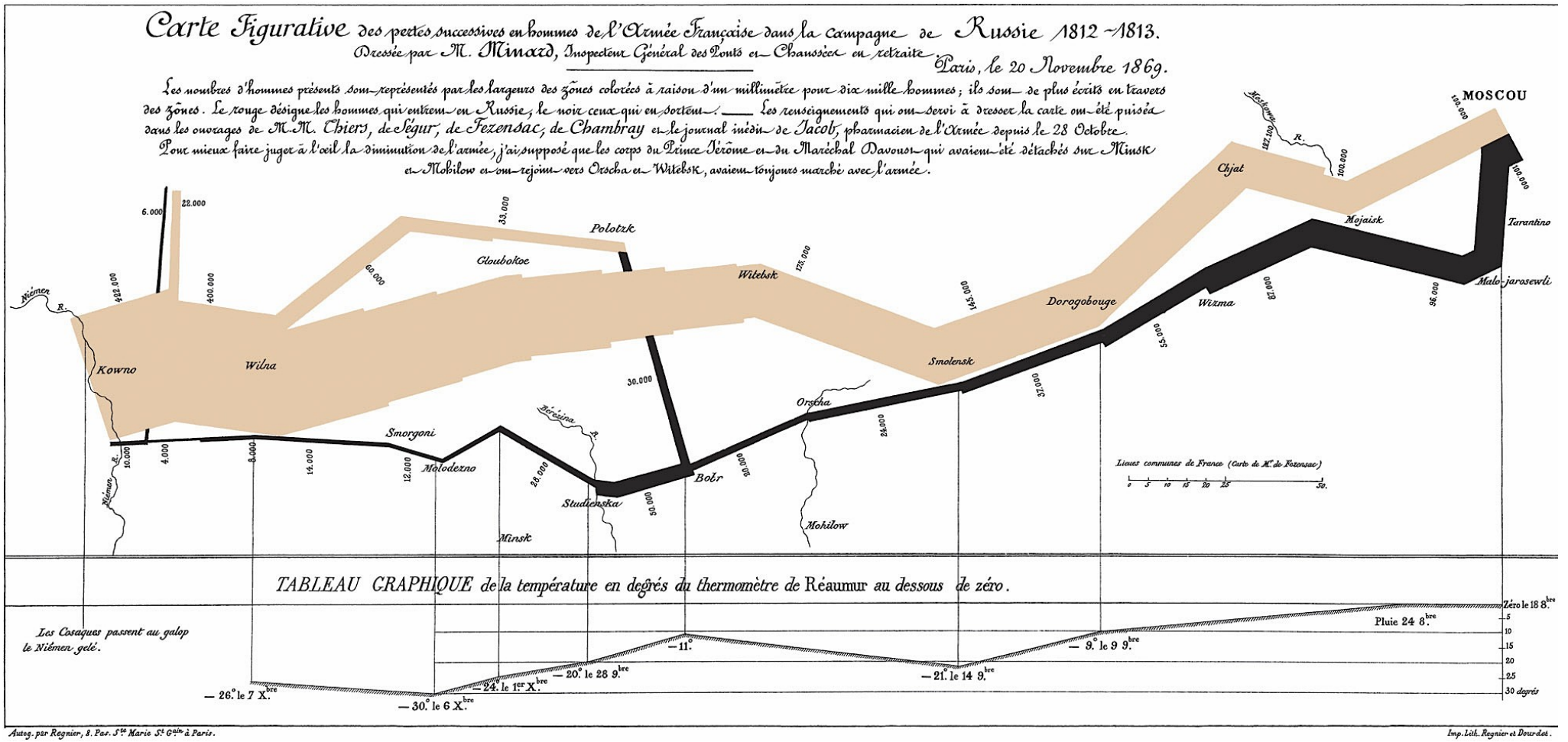
TYPE OF DISEASE

- Cancer
- Bone
- Skeletal
- Connective tissue
- Muscular
- Dermatological
- Cardiovascular
- Hematological
- Renal
- Endocrine
- Immunological
- Gastrointestinal
- Nutritional
- Metabolic
- Ear, nose, throat
- Ophthalmological
- Respiratory
- Developmental
- Neurological
- Psychiatric
- Multiple types
- Unclassified

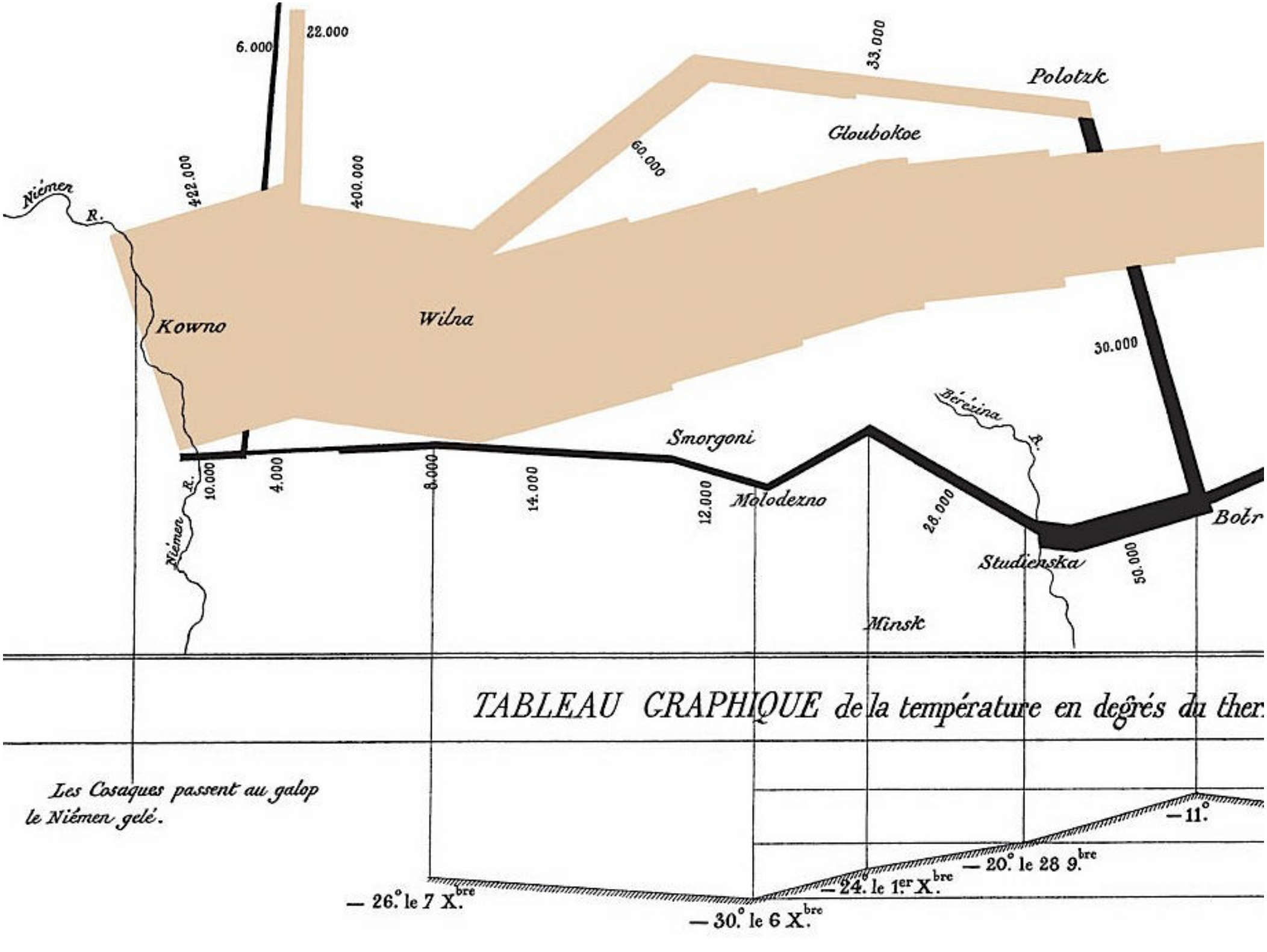
SCALE

Each circle represents a disease or disorder and is scaled in proportion to the number of genes associated with that disease.

Supergráficos



Marcha de Napoleão para Moscou na Guerra de 1812 (Charles Joseph Minard)

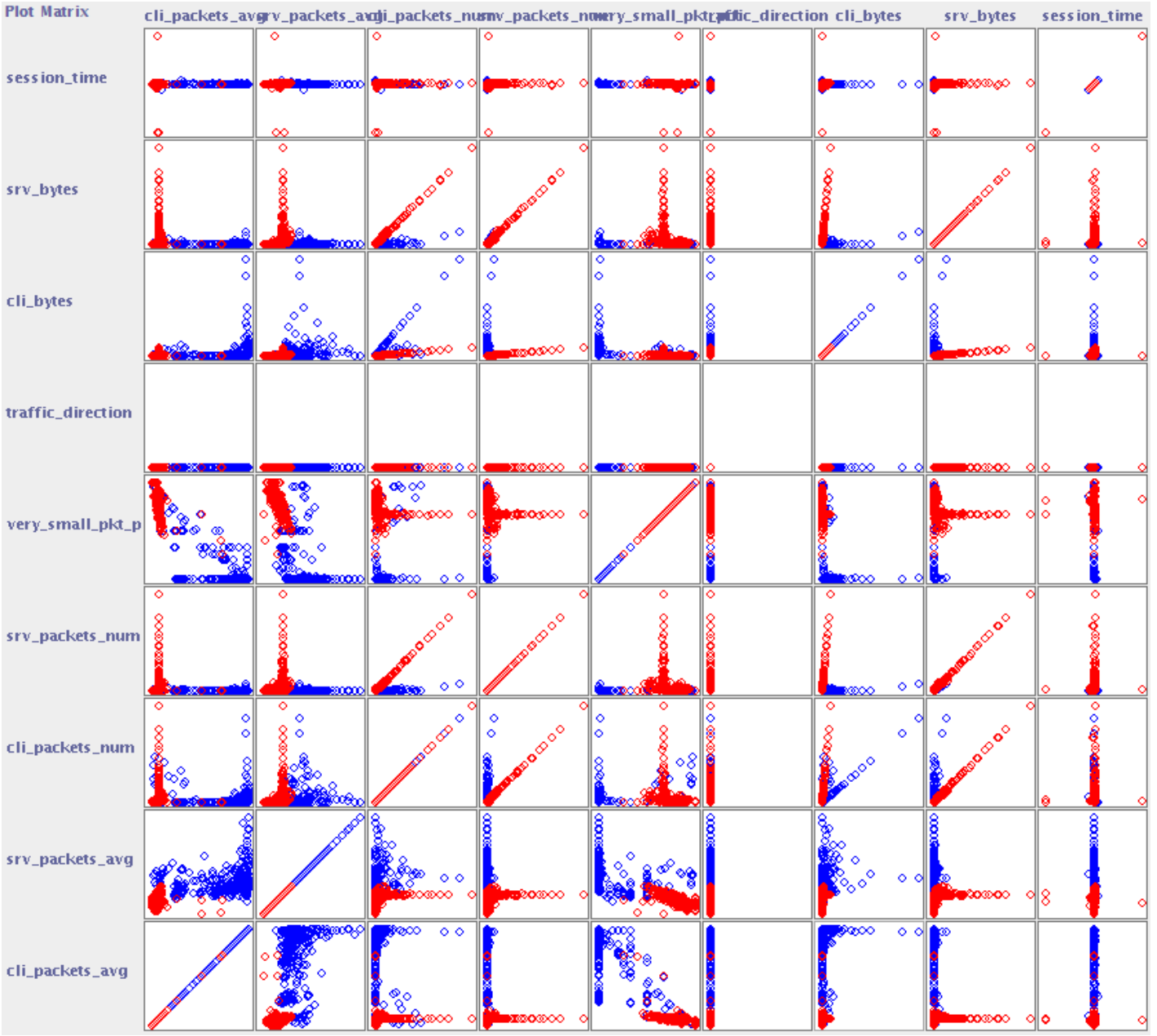

6.000
22.000
33.000
Polotzk
Glouboko̊e
60.000
Niémen R.
422.000
400.000
Kowno
Wilna
30.000
Bérézina R.
Smorgoni
Niémen R.
10.000
4.000
8.000
14.000
12.000
Molodezno
28.000
Bobr
Studienska
50.000
Minsk
TABLEAU GRAPHIQUE de la température en degrés du ther
Les Cosaques passent au galop le Niémen gelé.
— 11.°
— 20.° le 28 9.bre
— 24. le 1.er X.bre
— 26.° le 7 X.bre
— 30.° le 6 X.bre

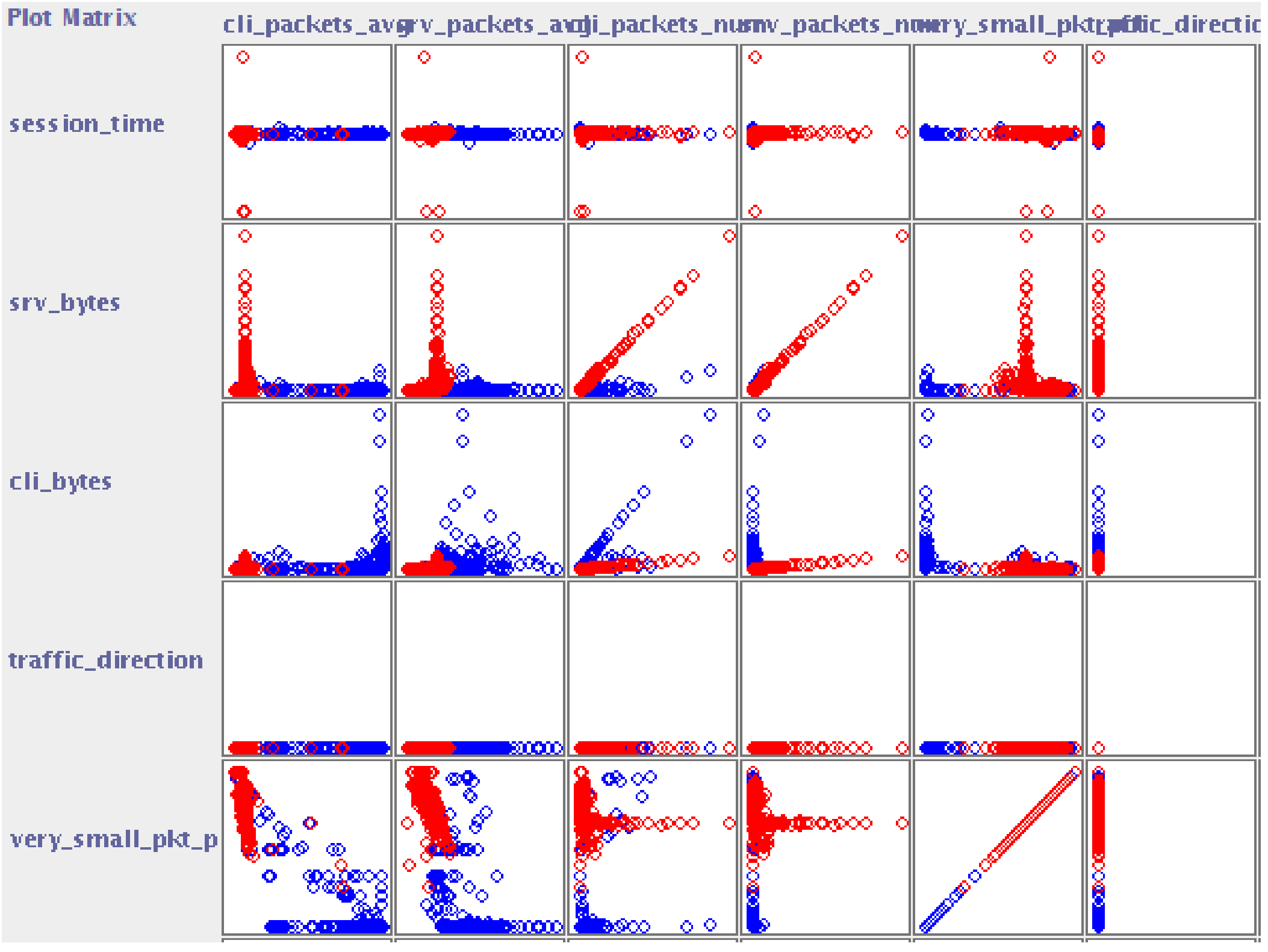
Subjetividade em Visualização

Subjetividade em Visualização

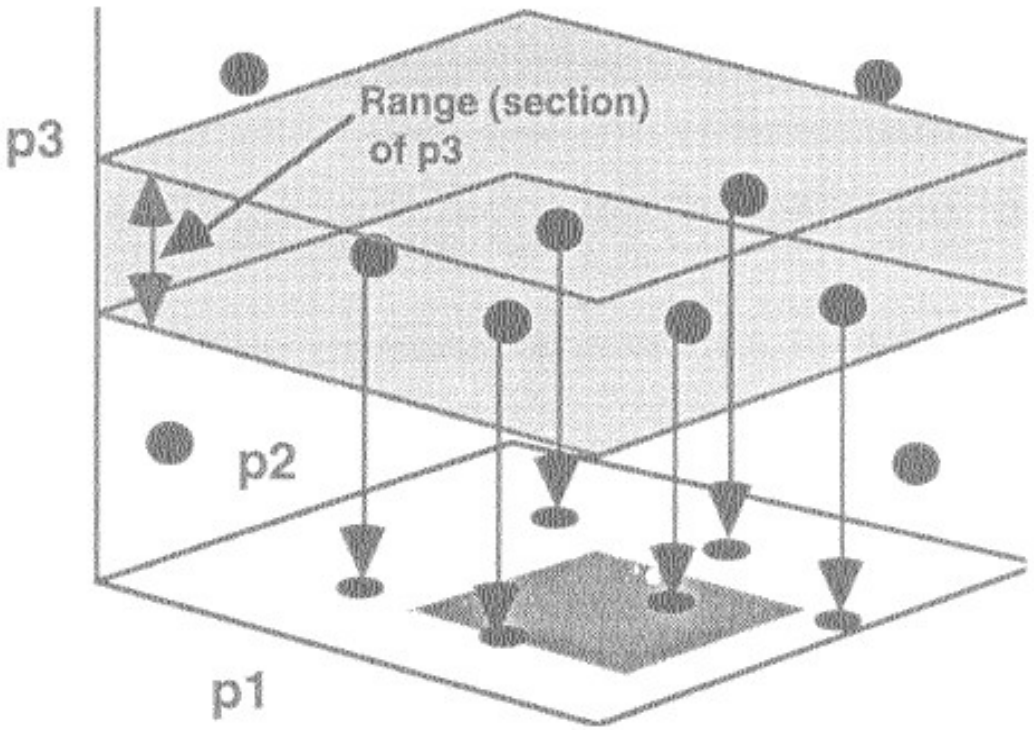
- **Técnicas Geométricas**
- Idéia básica: transformações e projeções usando arranjos em um número menor de dimensões.
  - *Scatterplot Matrices*: *K* atributos em grade *KxK*.
  - *Prosection Views*: *Scatterplot Matrices* com mecanismos de seleção (*drill-down*).
  - *Parallel Coordinates*: muito bom para dados mistos, requer exploração e rearranjos.
  - Visualização com Mapas de Kohonen (*SOMs*).

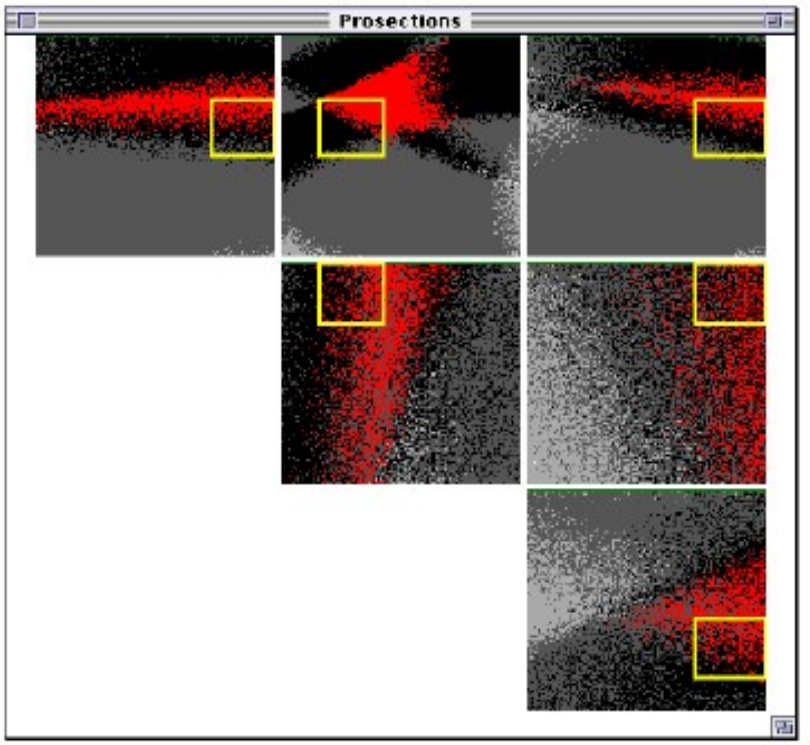
Técnicas de Visualização: *Scatterplot Matrices*




Plot Matrix
cli_packets_avg
srv_packets_avg
cli_packets_num
srv_packets_num
very_small_pkt
traffic_directio
session_time
srv_bytes
cli_bytes
traffic_direction
very_small_pkt_p

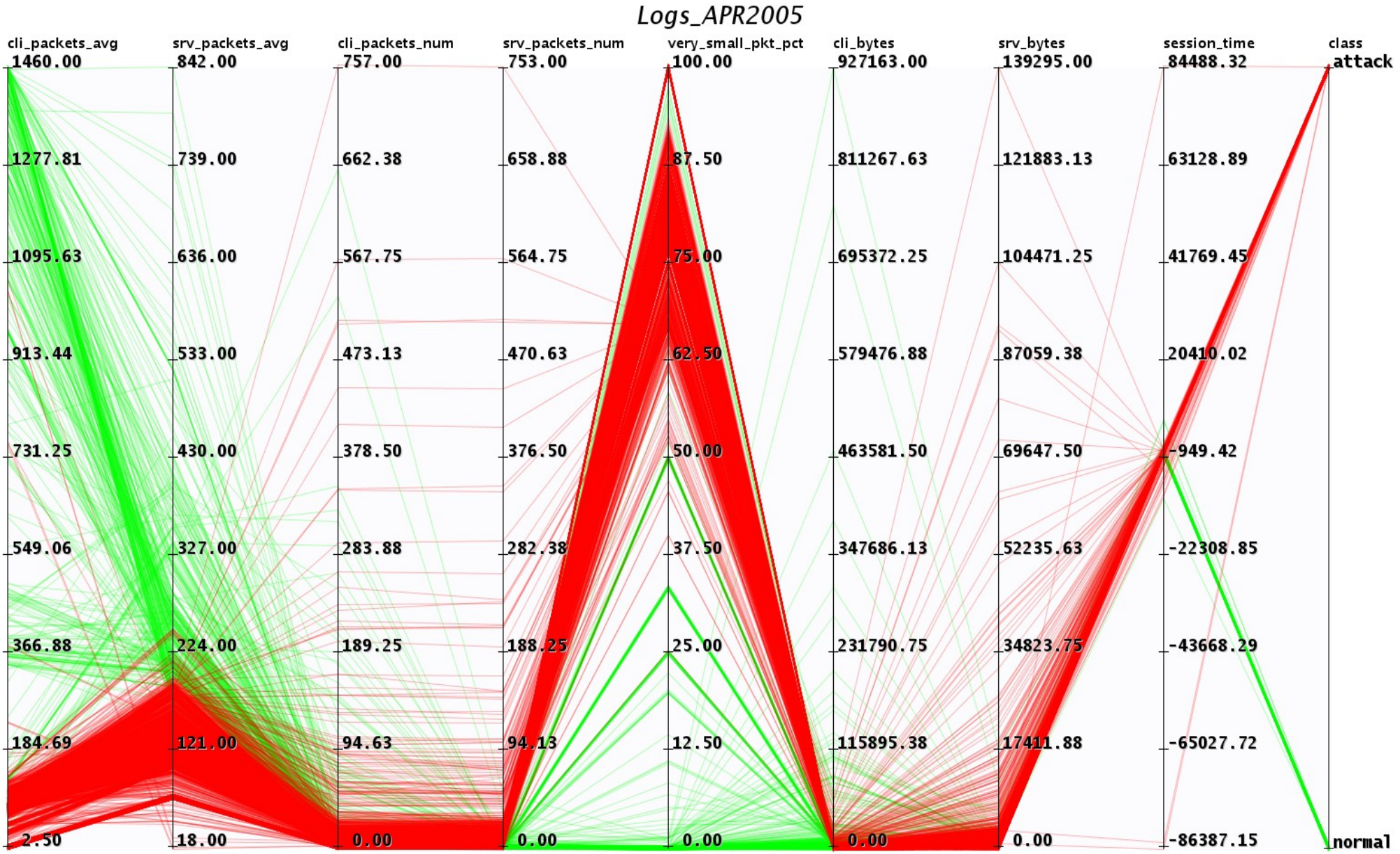
Técnicas de Visualização: *Prosection Views*



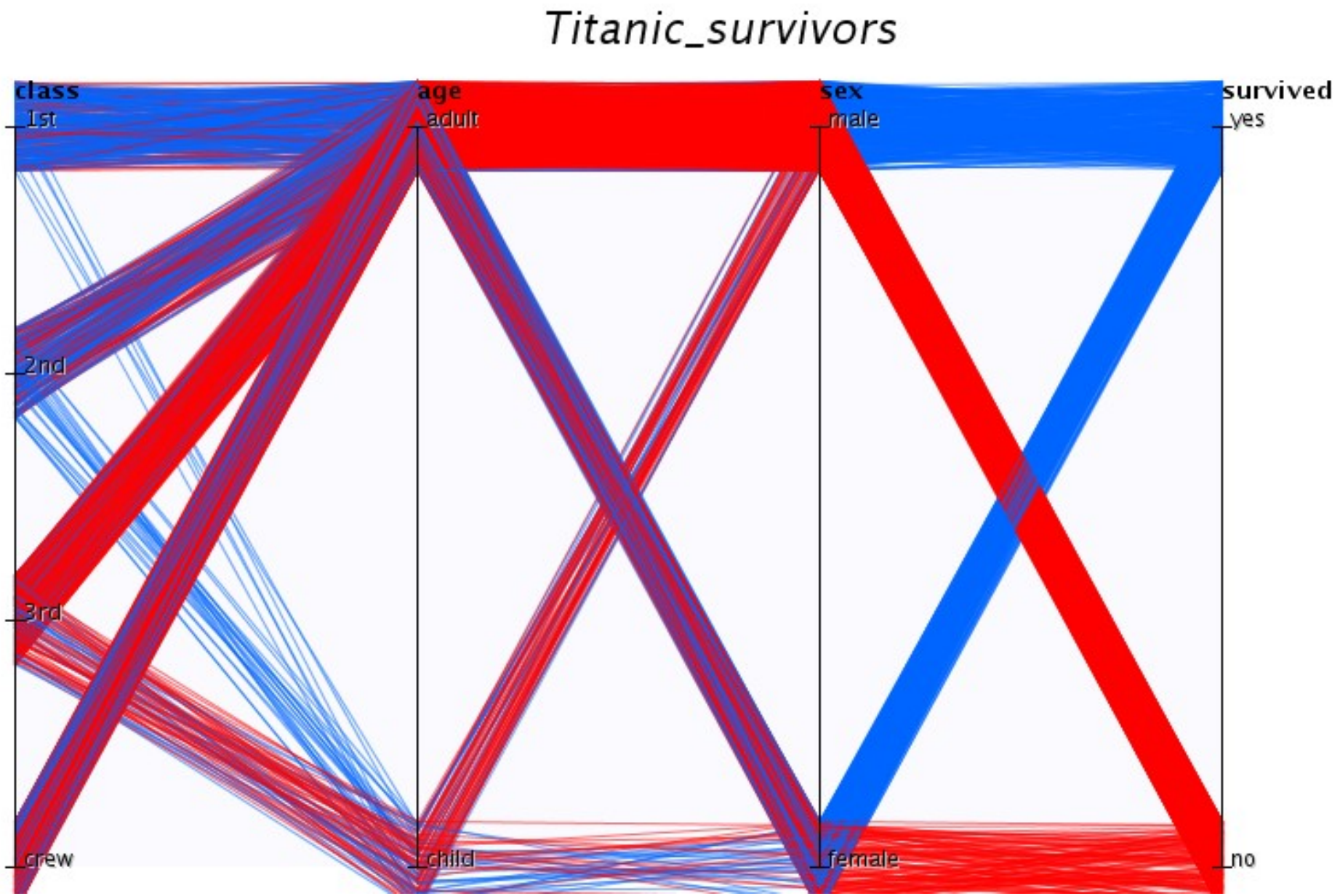


Exemplo de R. Spence, ilustrado no tutorial de Daniel Keim.

Técnicas de Visualização: *Parallel Coordinates*

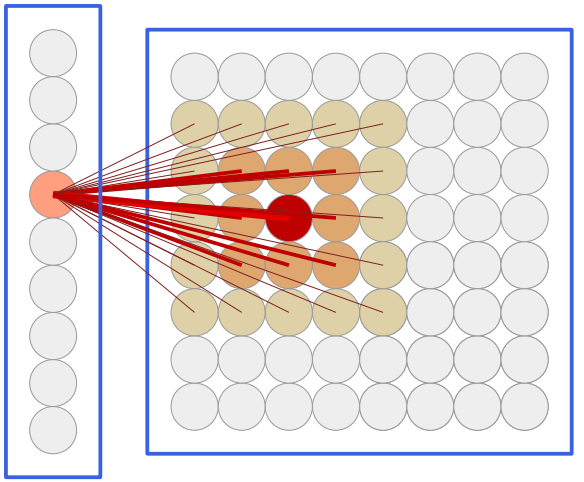


Técnicas de Visualização: *Parallel Coordinates*

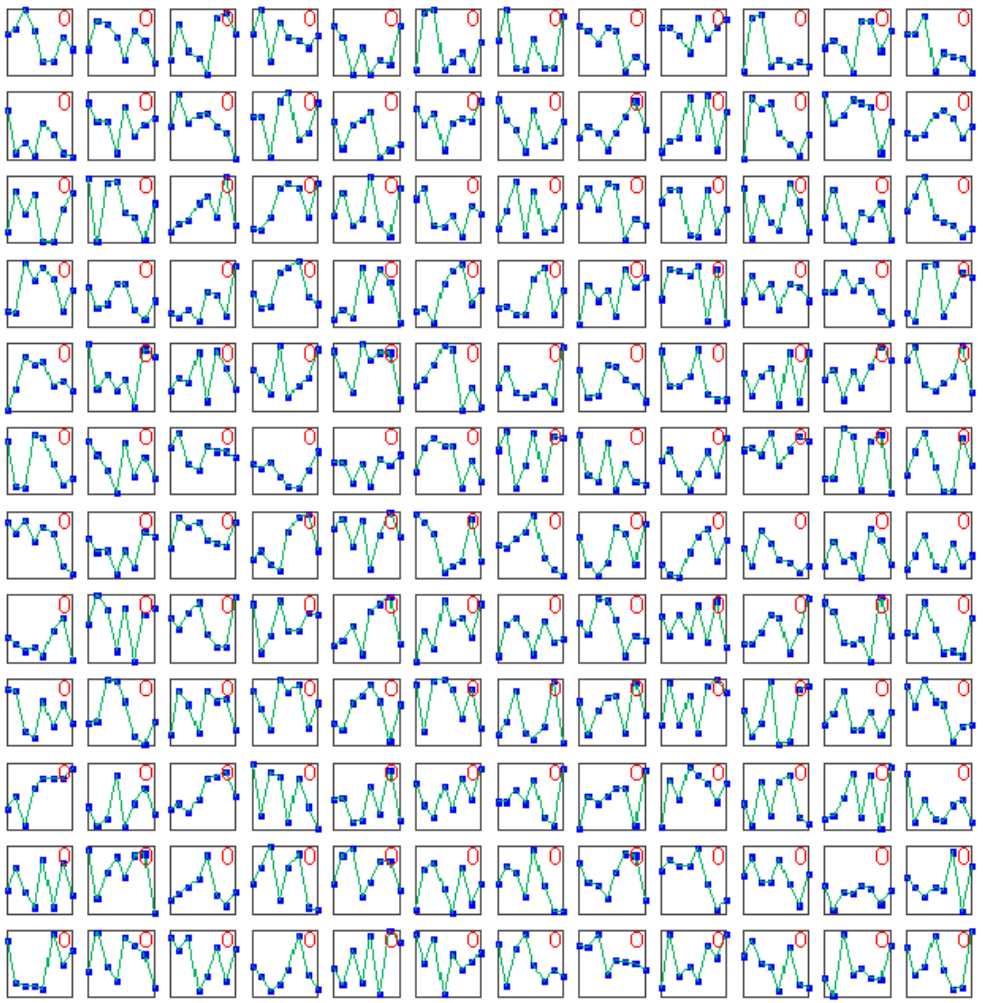


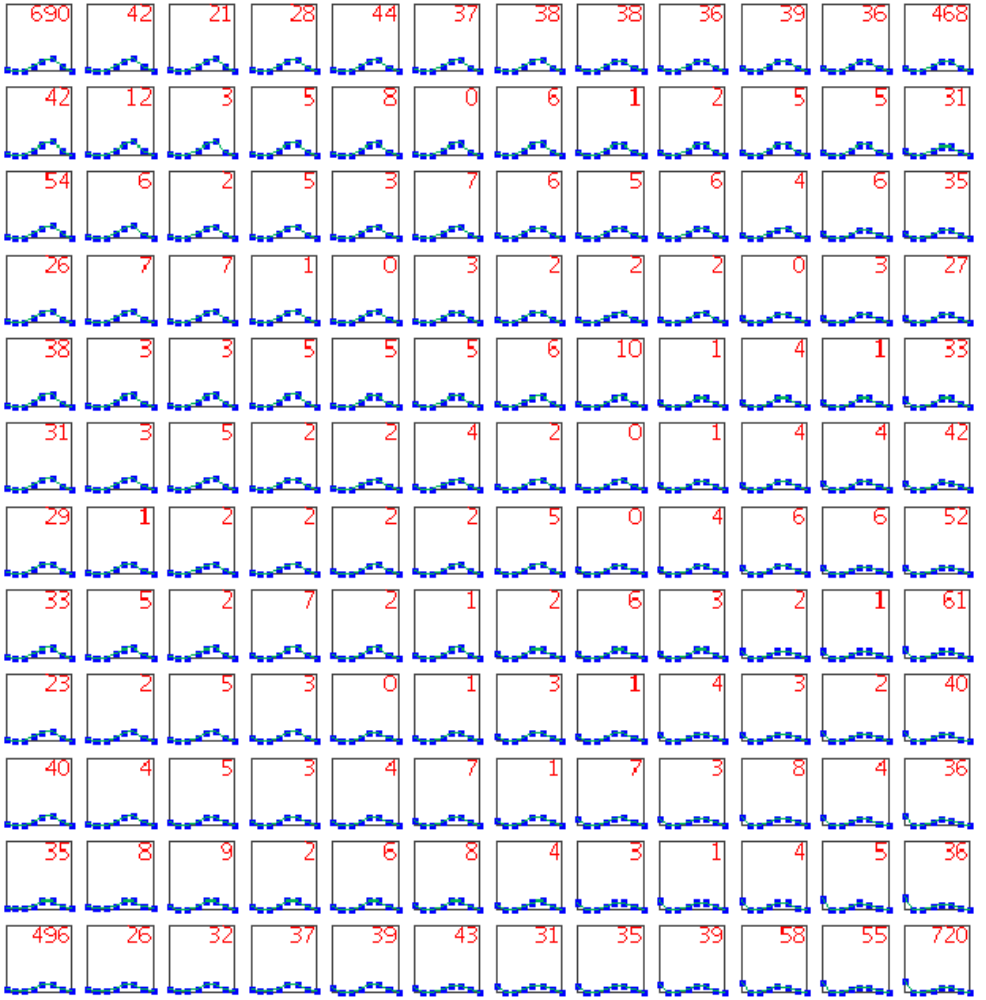
Técnicas de Visualização: *Self-Organizing Maps*

- Também conhecidos como redes de Kohonen.
  - Mapeiam vetores em N dimensões para 2 ou 3 dimensões, preservando topologia.
  - Por extensão, usados para fazer agrupamento e classificação em fase posterior.
  - Entrada: Vetores de dados, rede (considerar arquitetura), parâmetros de treinamento.
  - Saída: rede treinada, neurônios se assemelham a vetores apresentados.

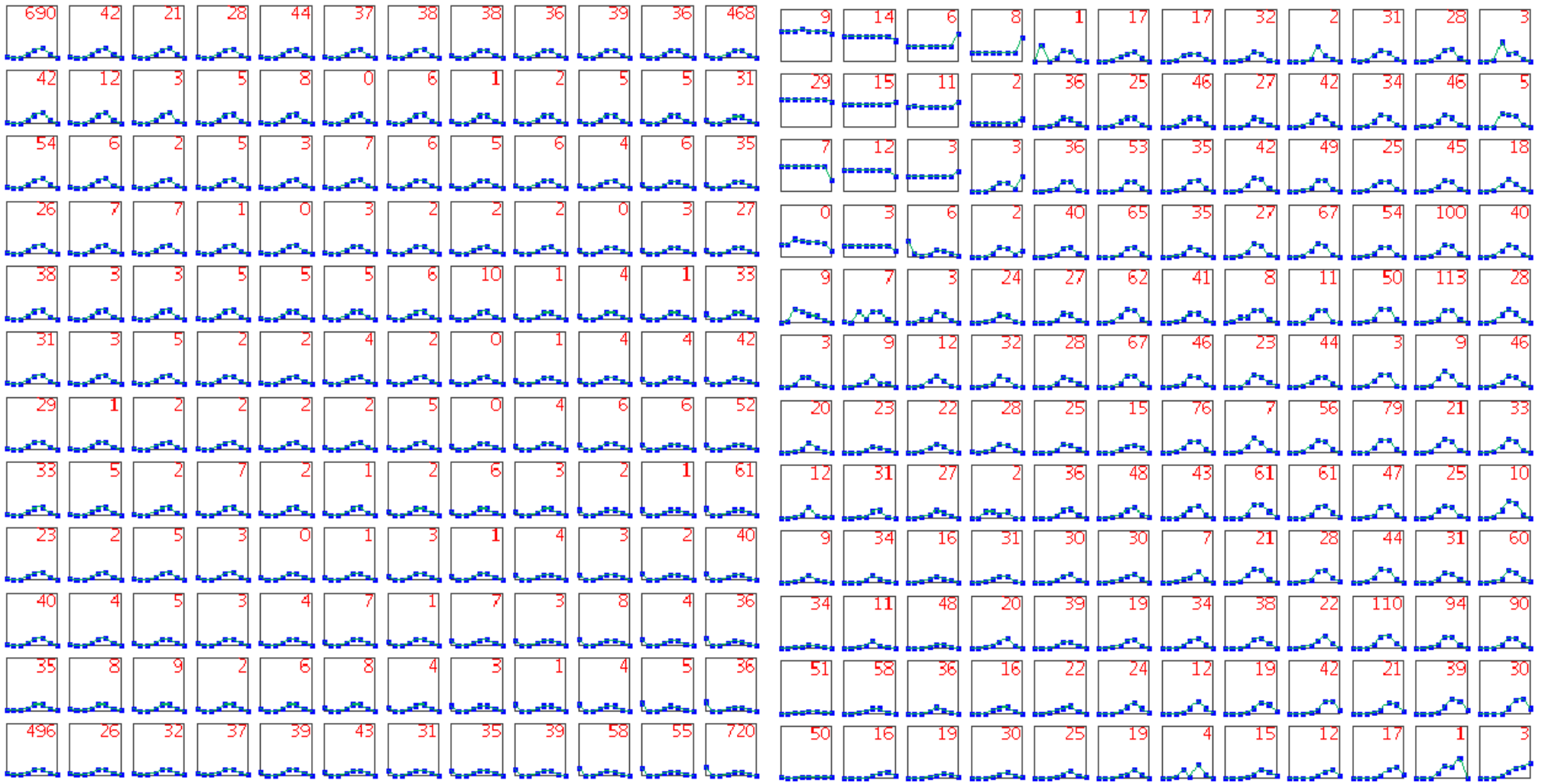


Técnicas de Visualização: *Self-Organizing Maps*

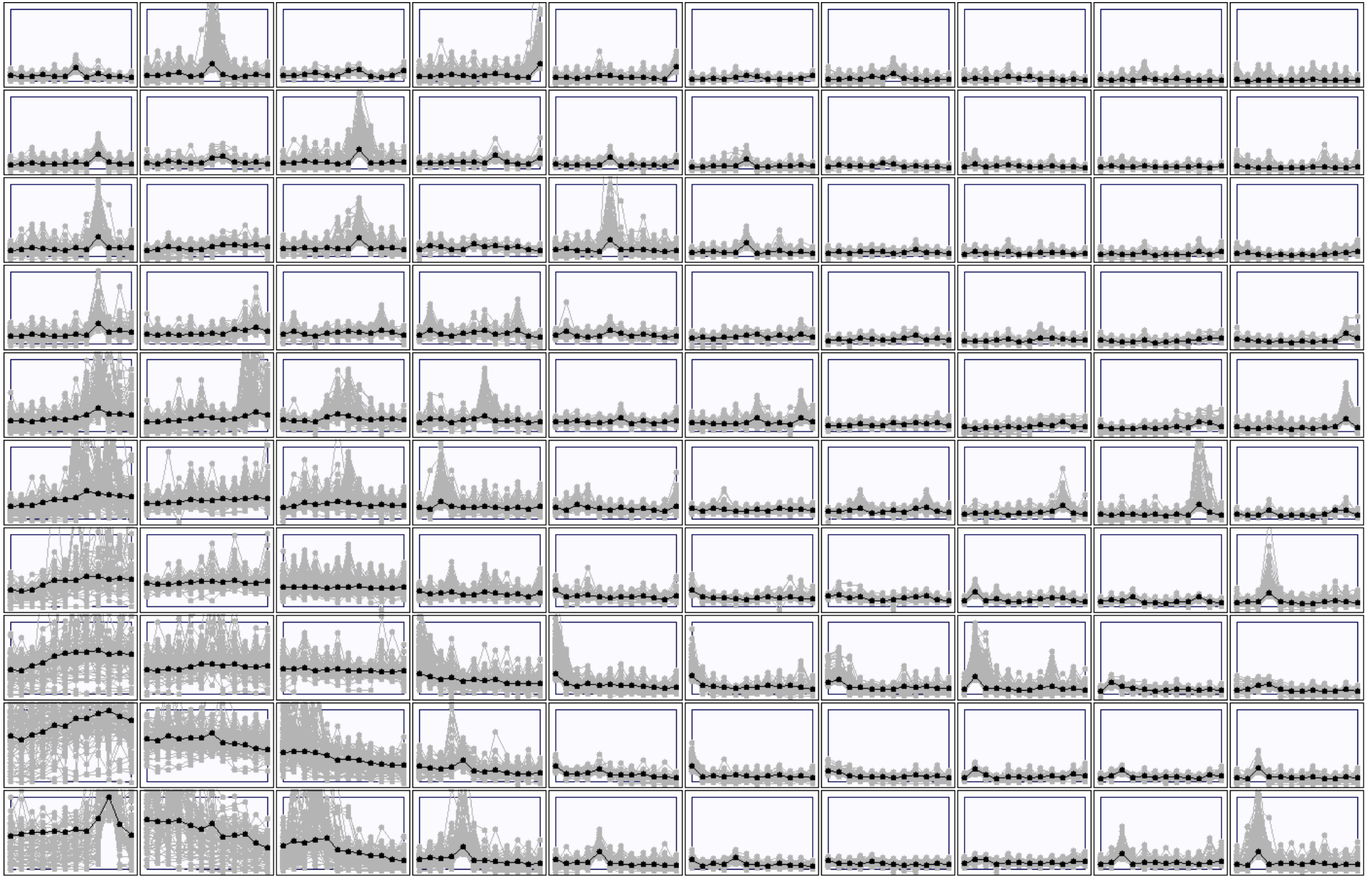




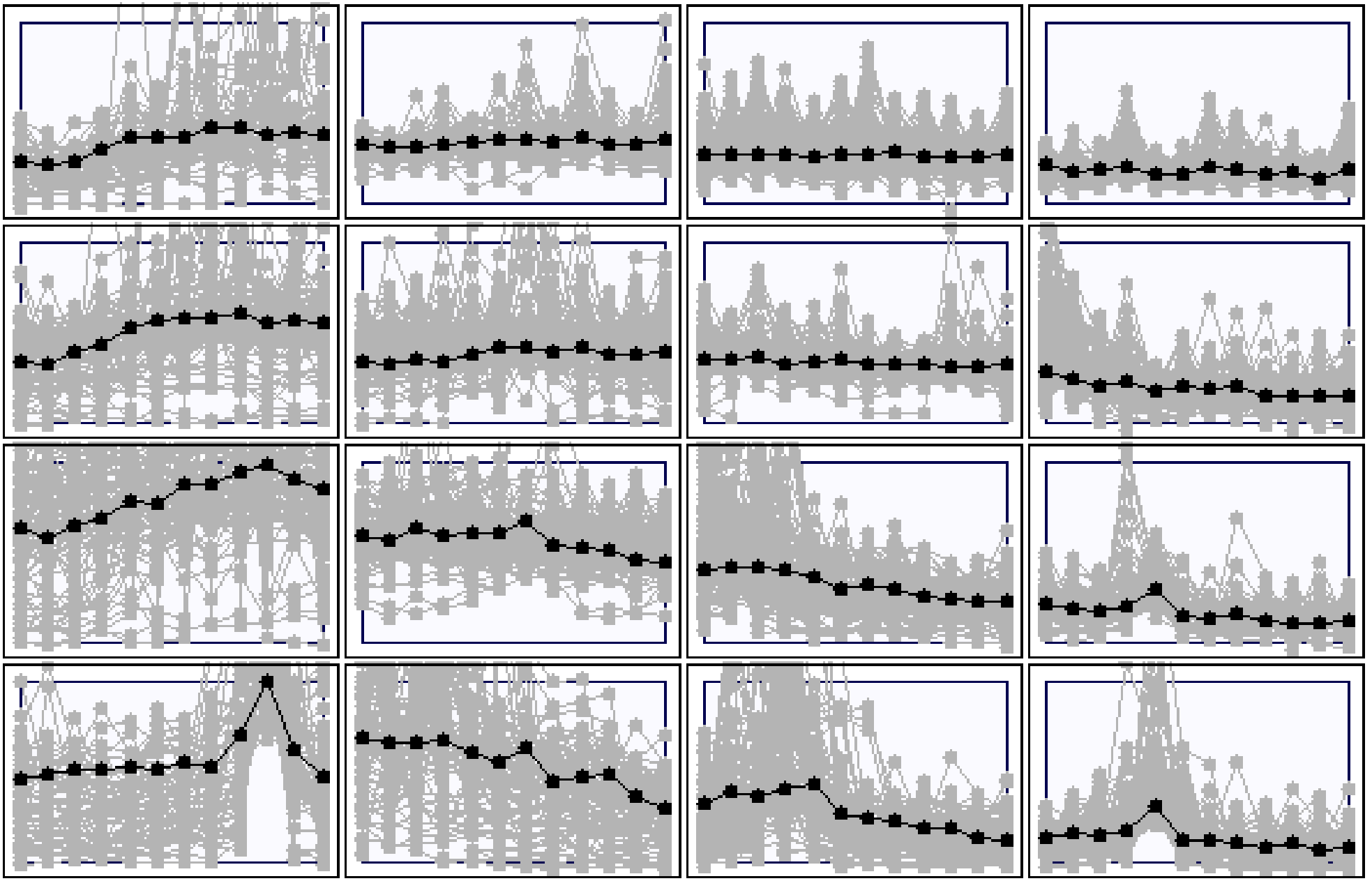
Técnicas de Visualização: *Self-Organizing Maps*



Técnicas de Visualização: *Self-Organizing Maps*

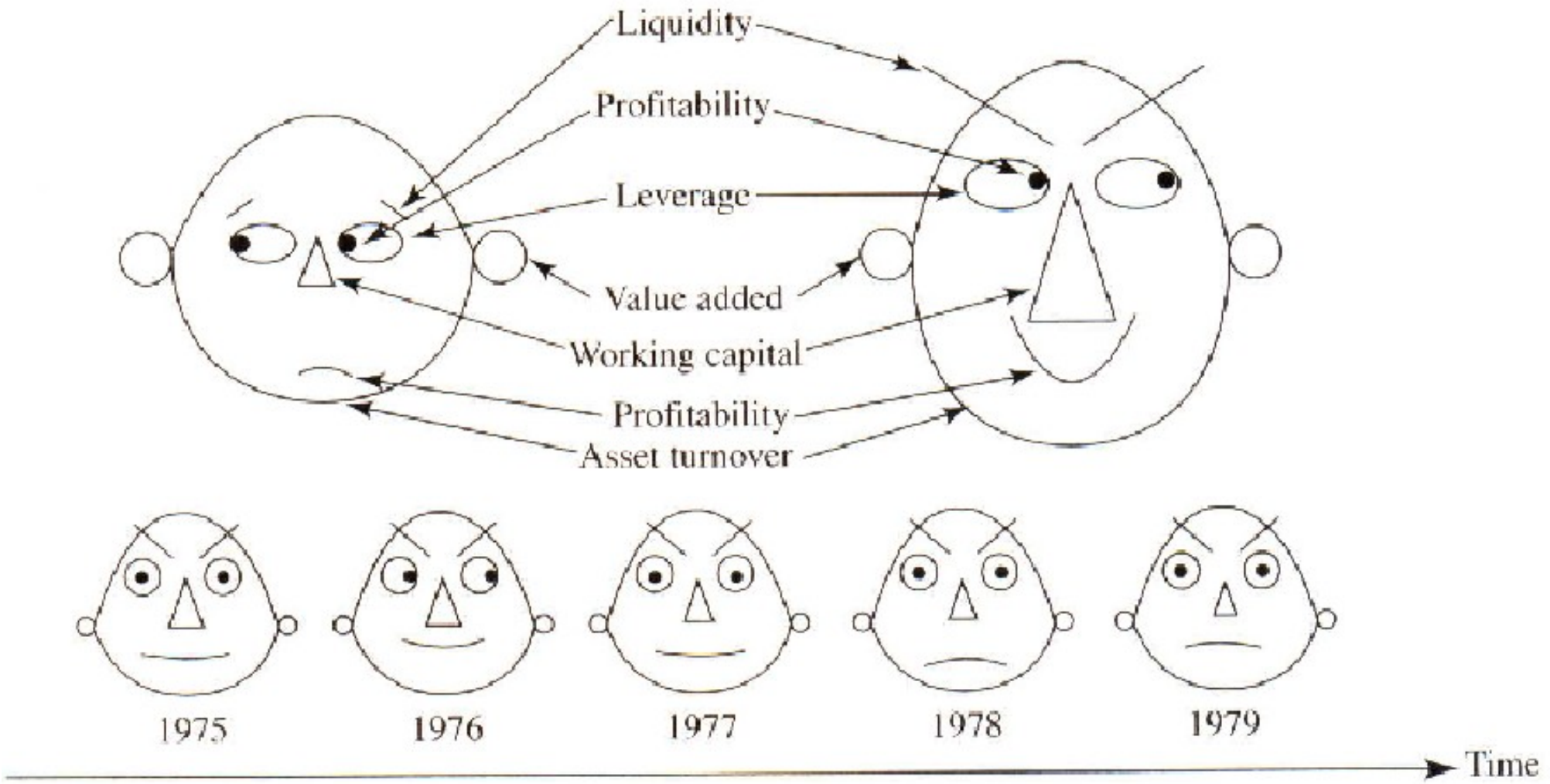


Técnicas de Visualização: *Self-Organizing Maps*

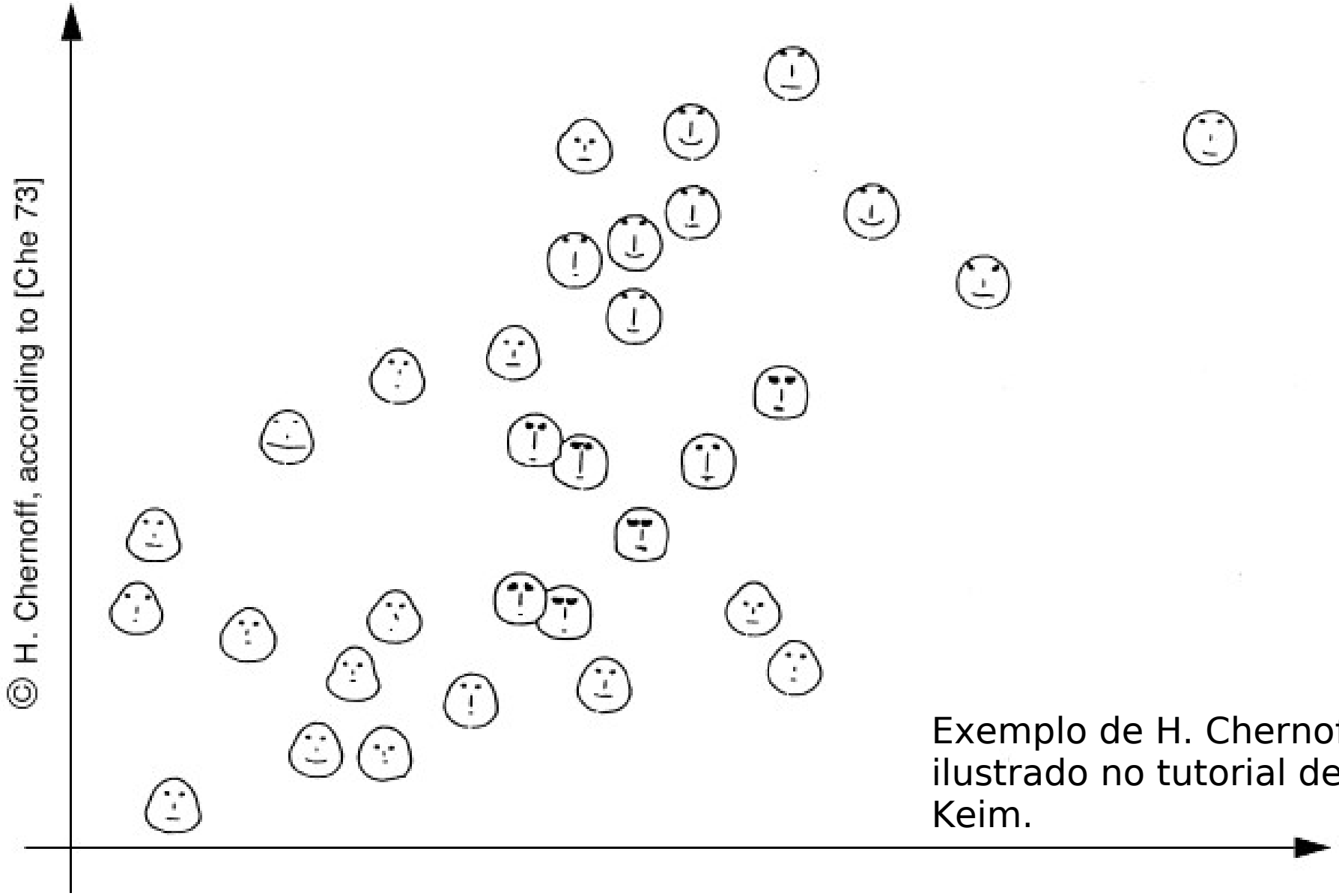


- **Técnicas Baseadas em Ícones**
- Idéia básica: usamos duas dimensões para mostrar ícones que representam outras dimensões adicionais.
  - Interpretação deve ser feita com legendas!
  - *Chernoff faces*: atributos das faces (geometria, olhos, excentricidade, curvaturas, etc.) representam outras dimensões.
  - *Stick figures*: dimensões adicionais mapeadas para ângulos e comprimentos de segmentos de retas.

Técnicas de Visualização: *Chernoff Faces*

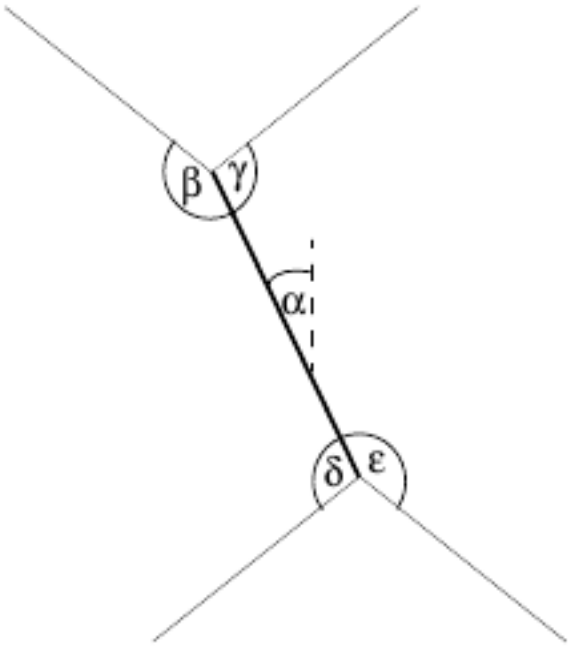


Técnicas de Visualização: *Chernoff Faces*

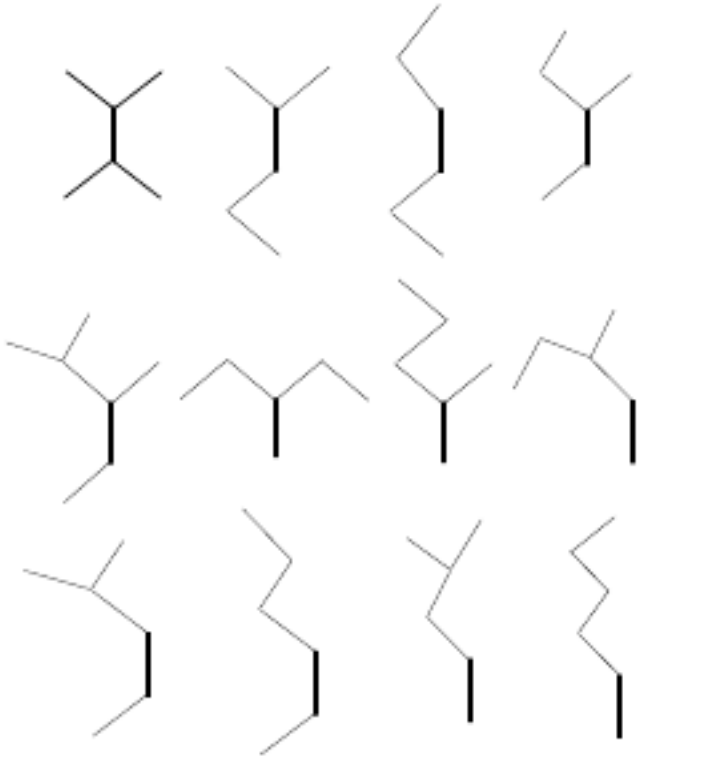


Exemplo de H. Chernoff, ilustrado no tutorial de Daniel Keim.

Técnicas de Visualização: *Stick Figures*



Stick Figure Icon



A Family of Stick Figures

- Uso de duas dimensões mais textura

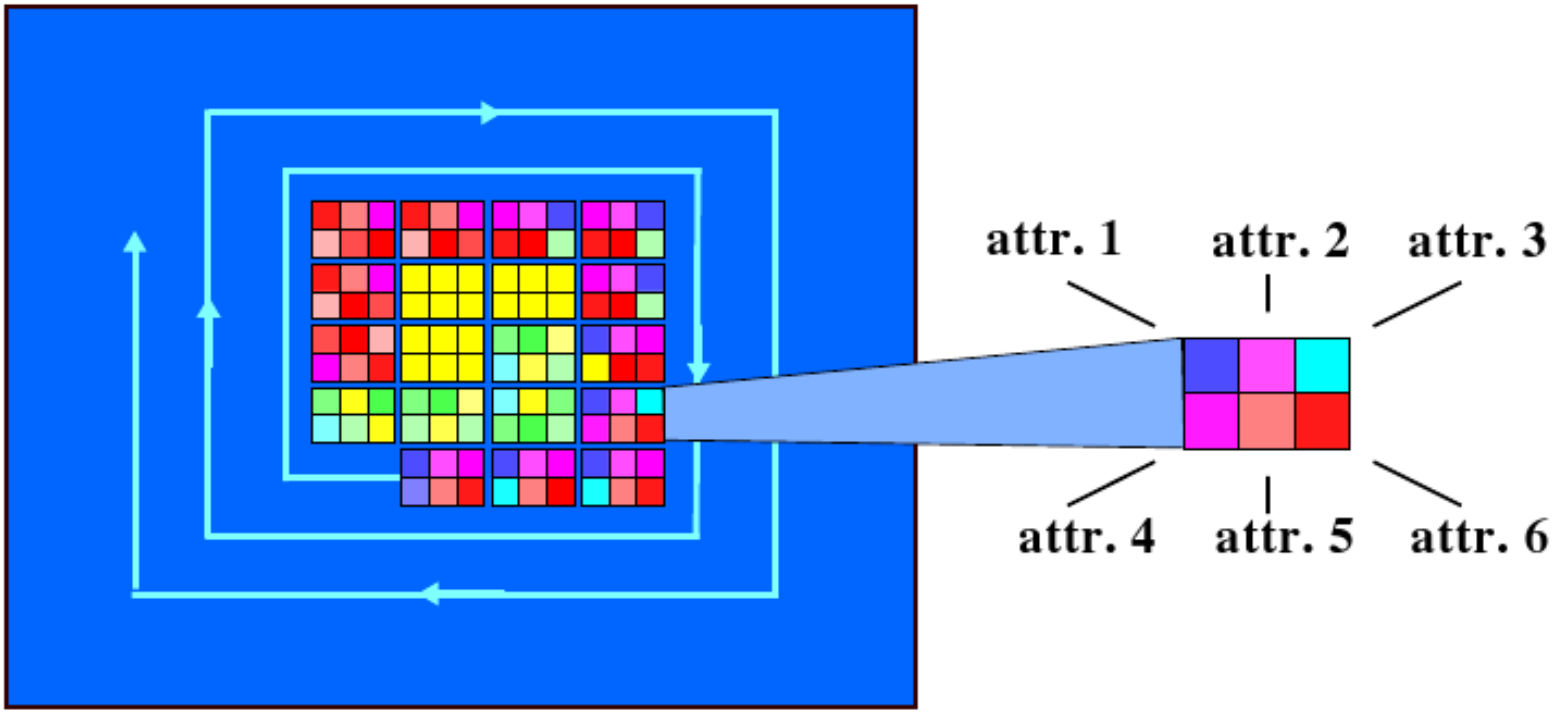
Fonte: Tutorial de Daniel Keim.

- **Técnicas Baseadas em Pixels**
- Idéia básica: ícones pequenos, uso de cores, geometria simples.
  - Interpretação mais instintiva, menos uso de legendas.
  - Distribui pixels em duas dimensões que podem ou não ser índices (podendo ou não causar artefatos!).
  - Existem várias maneiras de organizar pixels em duas dimensões.

Técnicas de Visualização: *Grouping Technique*

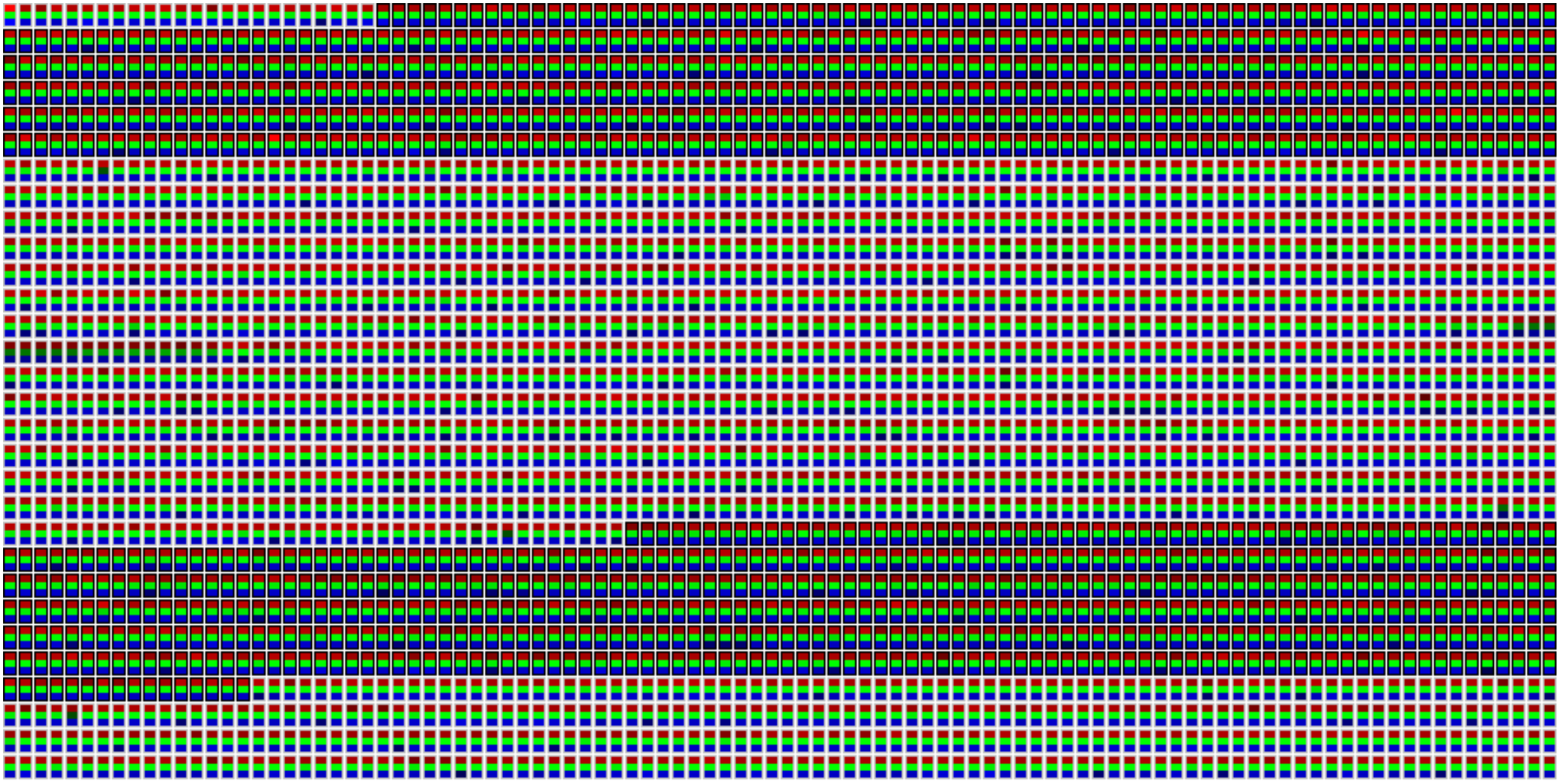
- ⇨ visualization of the data using color icons
- ⇨ color icons are array of color fields representing the attribute values
- ⇨ arrangement is query-dependent (e.g., spiral)

schematic representation of 6-dim. data



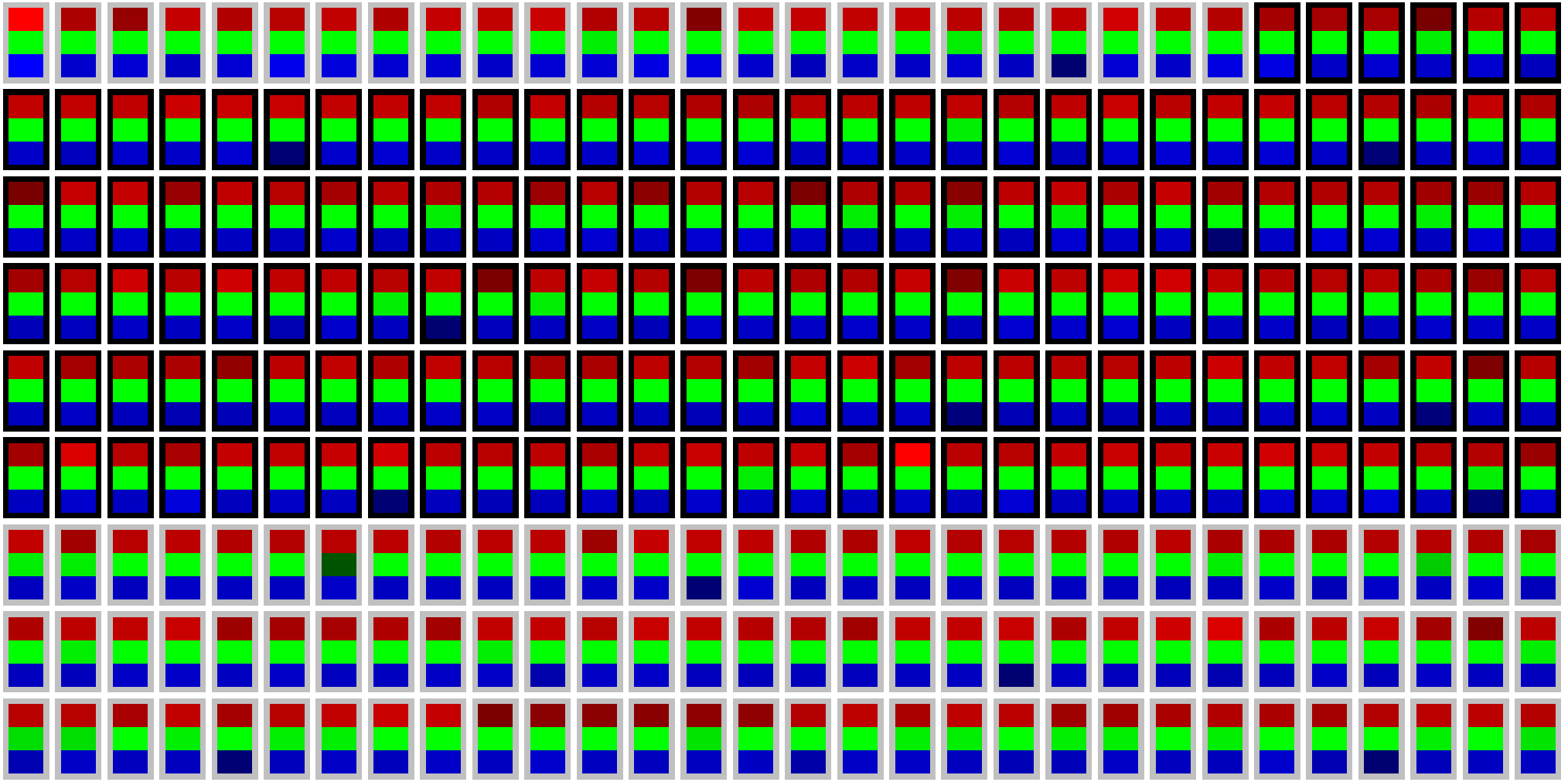
Fonte: Tutorial de Daniel Keim.

Técnicas de Visualização: *Grouping Technique*



Pacotes TCP, UDP e ICMP recebidos por honeypots em 10 dias (a cada 20 minutos).

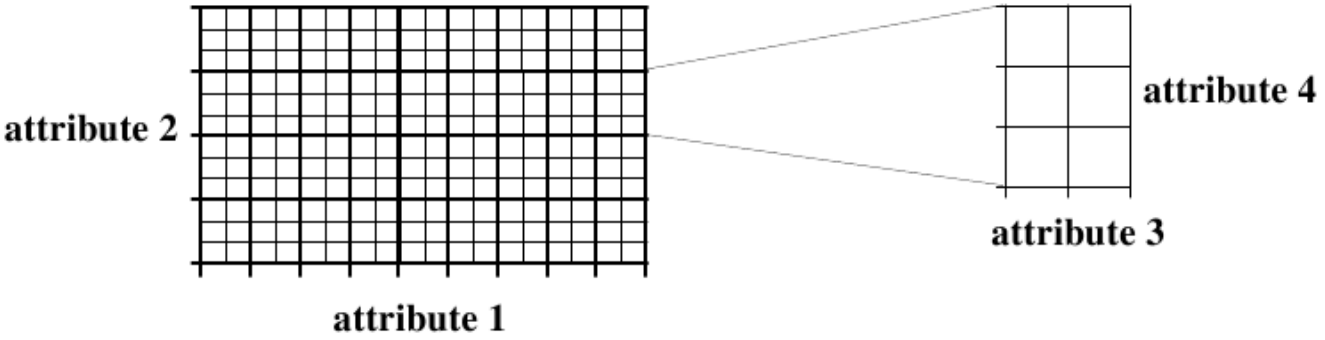
Técnicas de Visualização: *Grouping Technique*

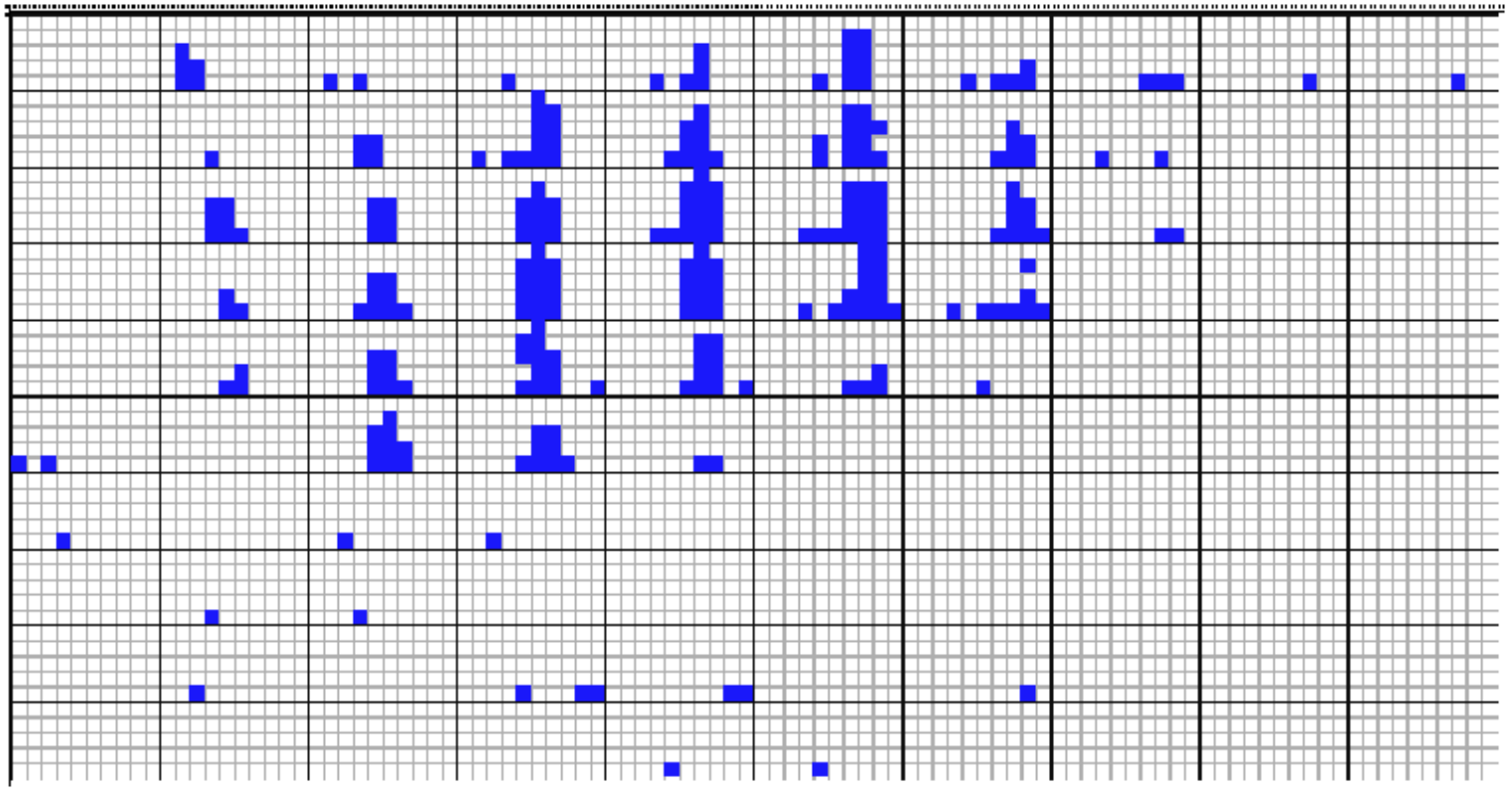


Pacotes TCP, UDP e ICMP recebidos por honeypots em 10 dias (a cada 20 minutos).

- **Técnicas Hierárquicas**
- Idéia básica: particionamento das dimensões em subdimensões.
  - *Dimensional Stacking:* Particionamento de N dimensões em conjuntos de 2 dimensões.
  - *Treemap:* Preenche área de visualização alternando eixos X e Y.
  - *Cone Trees:* Visualização interativa de dados hierárquicos.
  - *InfoCube:* Visualização hierárquica com 3D e transparência.

Técnicas de Visualização: *Dimensional Stacking*



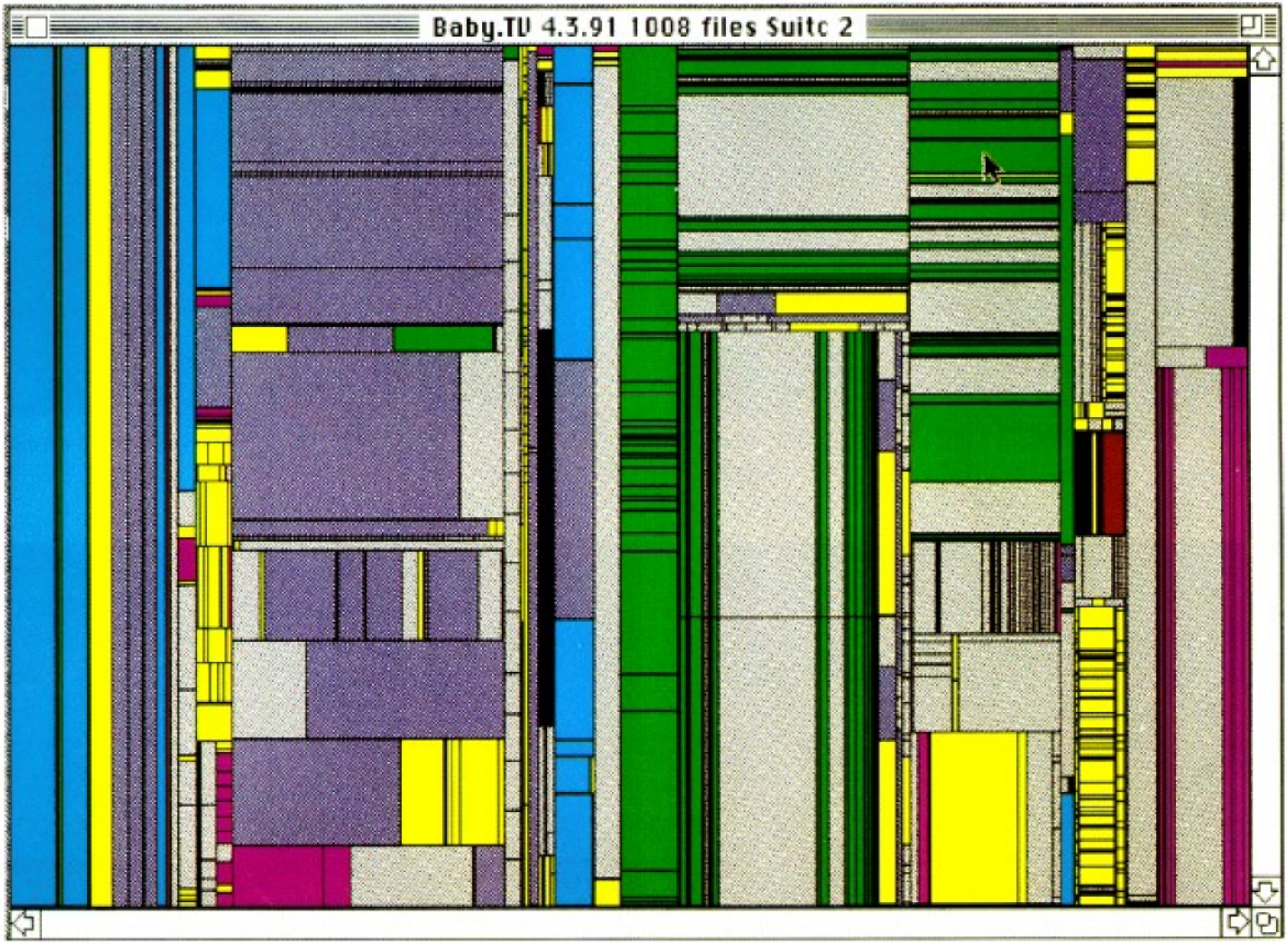


used by permission of M. Ward, Worcester Polytechnic Institute

visualization of oil mining data with longitude and latitude mapped to the outer x-, y- axes and ore grade and depth mapped to the inner x-, y- axes

Fonte: Tutorial de Daniel Keim.

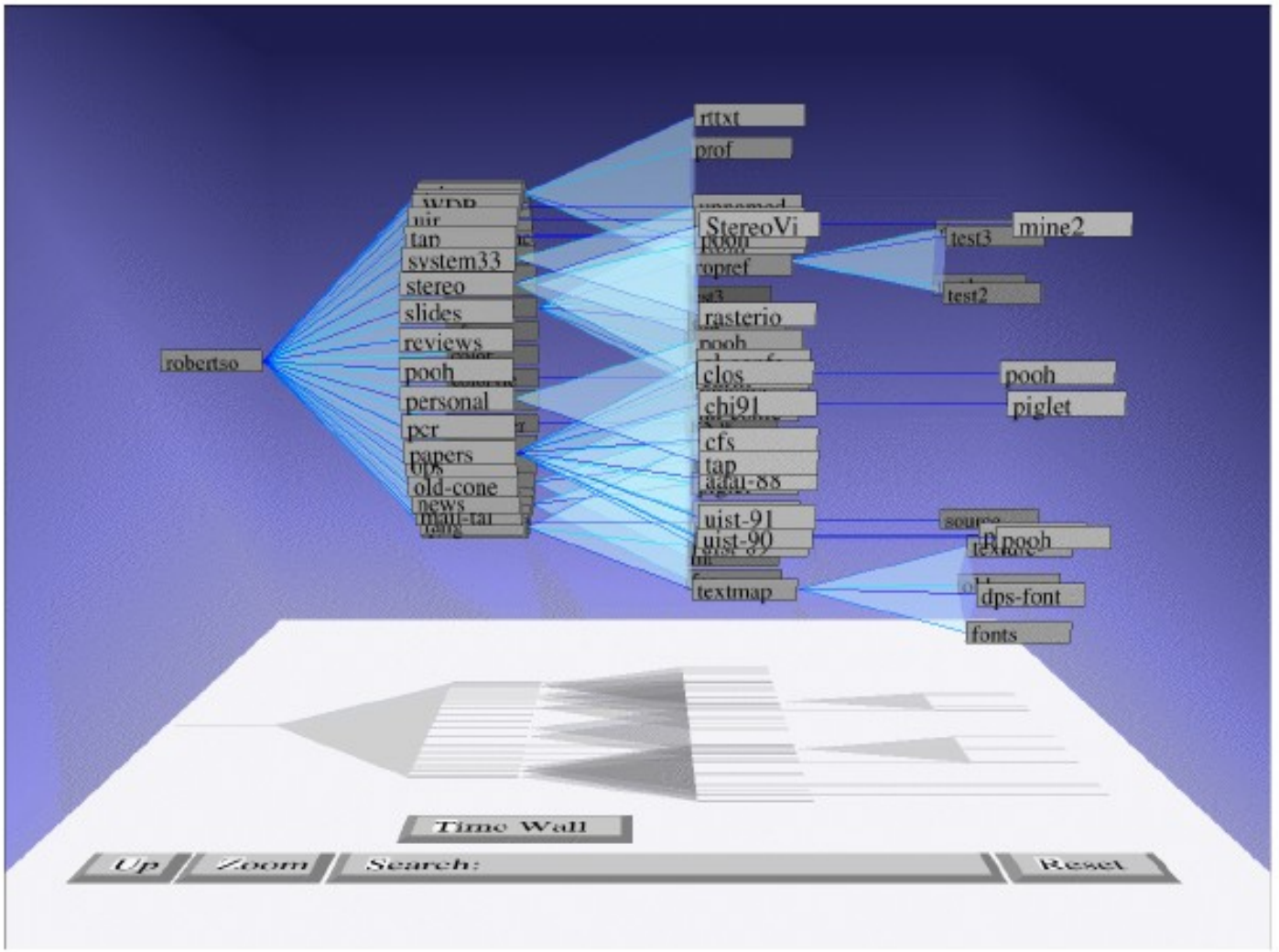
Técnicas de Visualização: *Treemap*



treemap of a file system containing about 1000 files

Fonte: Tutorial de Daniel Keim.

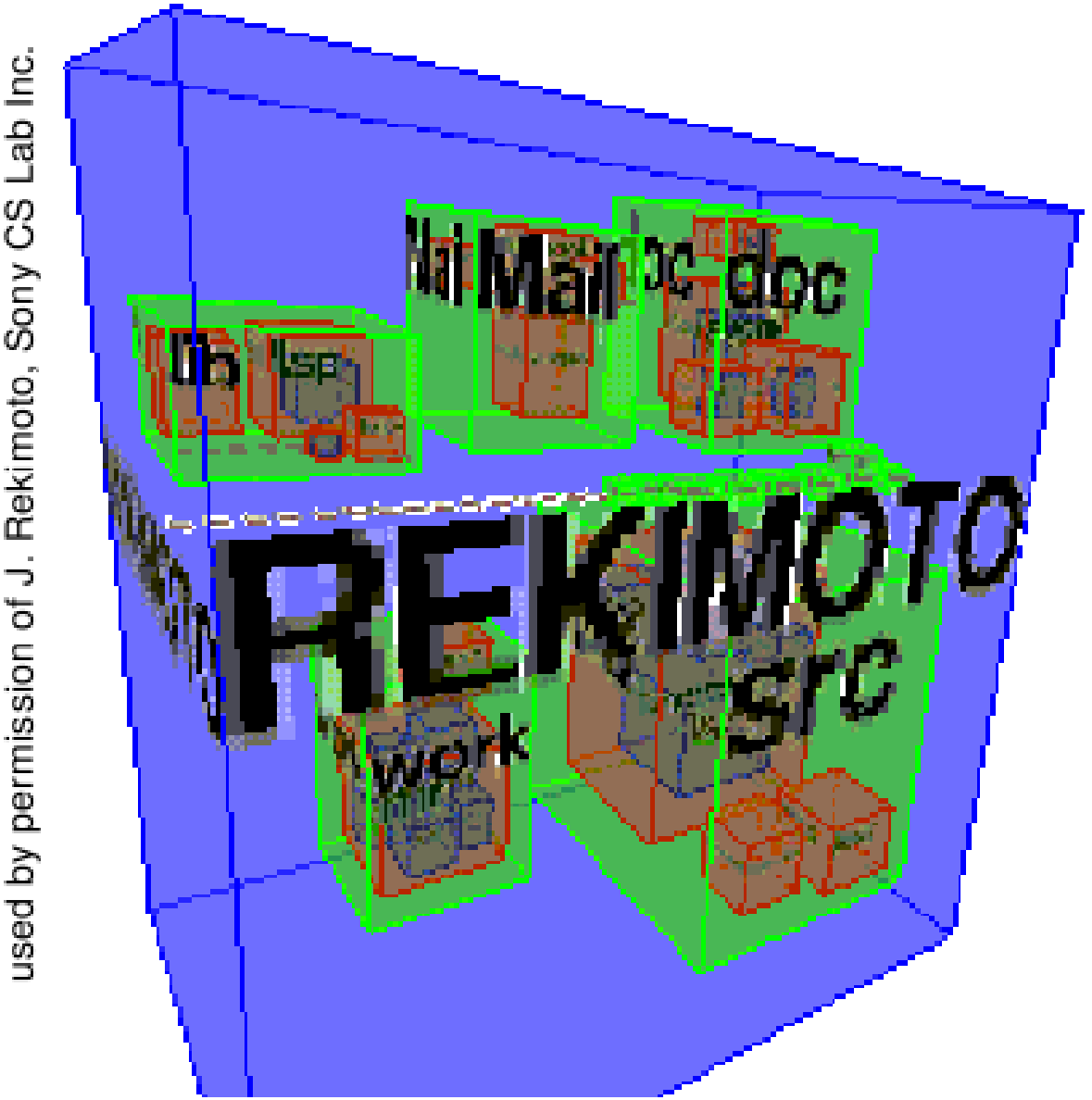
Técnicas de Visualização: *Cone Tree*



file system structure visualized as a cone tree

Fonte: Tutorial de Daniel Keim.

Técnicas de Visualização: *Info Cube*

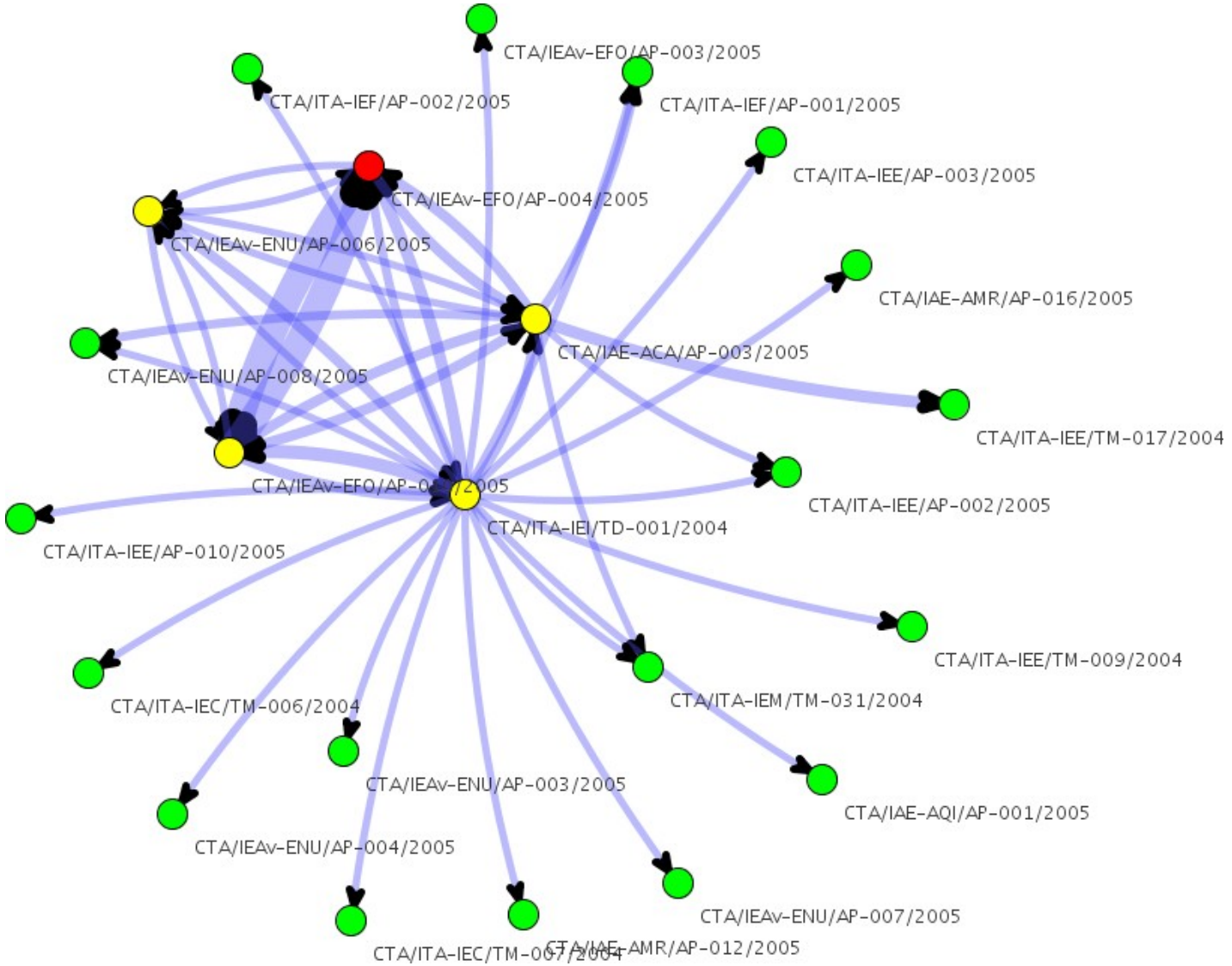


Fonte: Tutorial de Daniel Keim.

- **Técnicas Baseadas em Grafos**
- Idéia básica: conjunto de pontos (vértices) ligados por linhas (as arestas).
  - Representam conexões ou ligações de alguma forma.
  - Enorme variabilidade na organização geométrica dos vértices e arestas.
  - Representações gráficas diferentes para vértices e arestas.

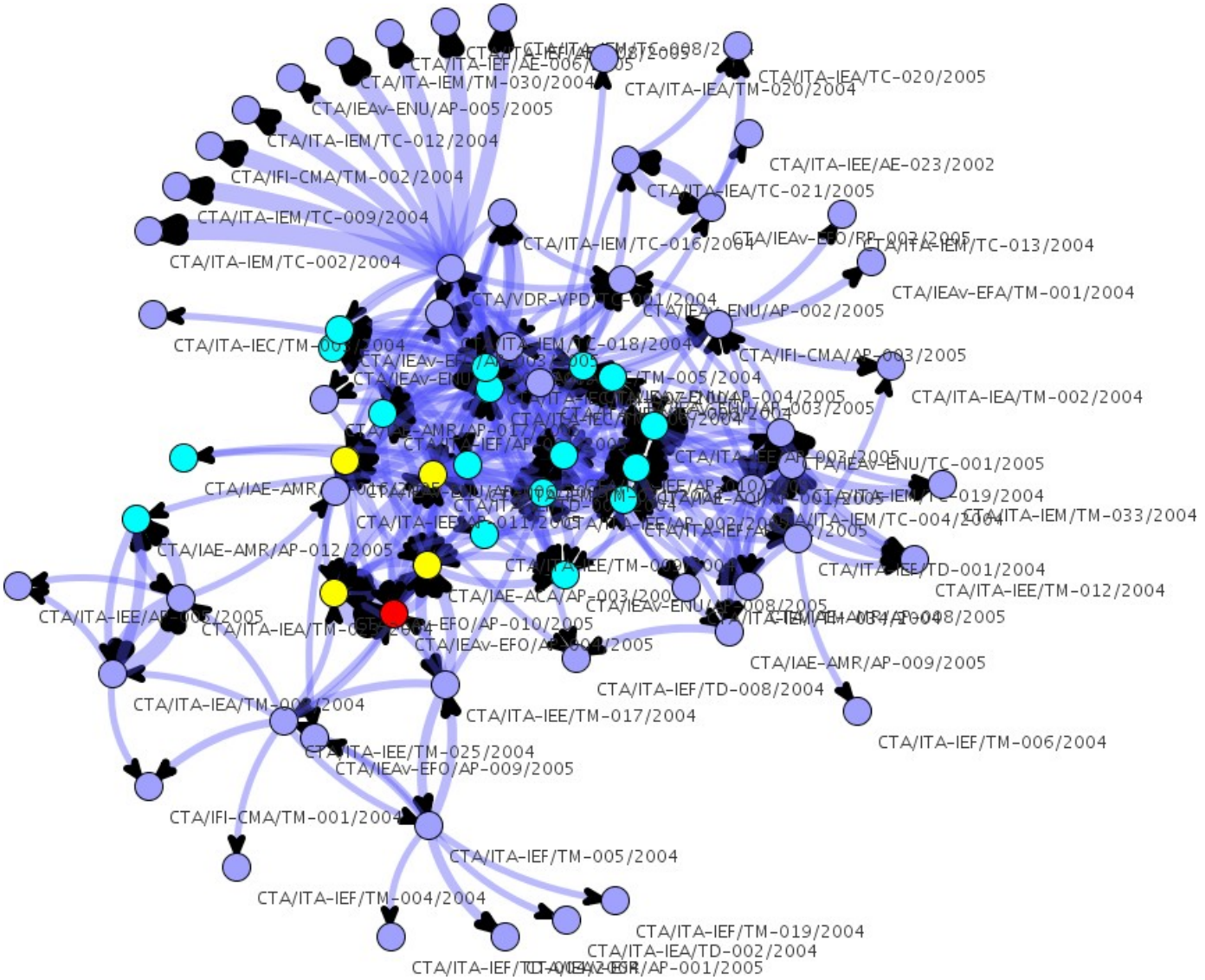
- Representação para visualização → *mineração de grafos*.

Técnicas de Visualização: Grafos



Um Sistema de Recomendação de Publicações Científicas Baseado em Avaliação de Conteúdo, Relatório Final de Alessandro Oliveira Arantes, disciplina CAP-359, INPE.

Técnicas de Visualização: Grafos

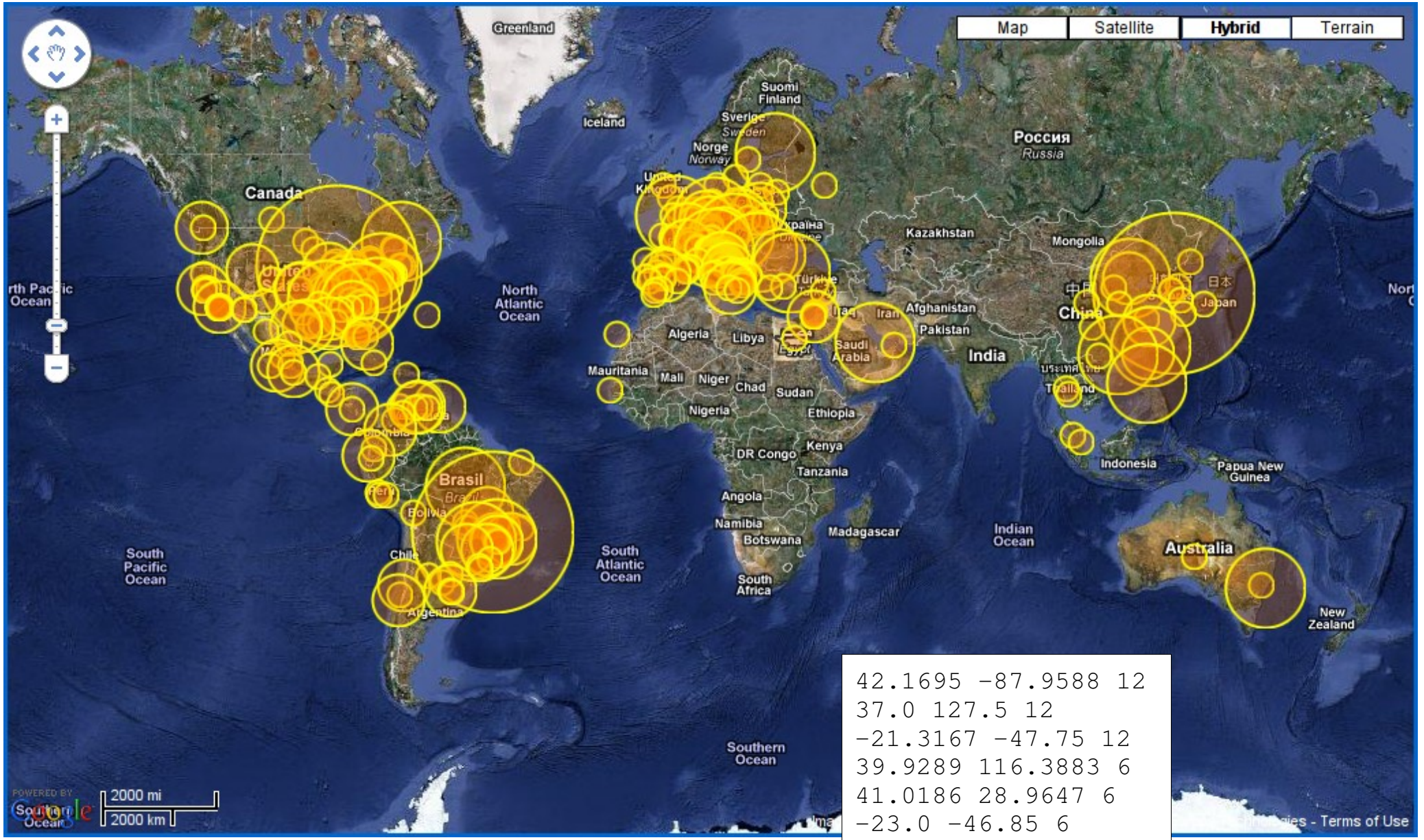


Um Sistema de Recomendação de Publicações Científicas Baseado em Avaliação de Conteúdo, Relatório Final de Alessandro Oliveira Arantes, disciplina CAP-359, INPE.

- **Técnicas Tridimensionais**
- Idéia básica: recursos de computação gráfica para usar dimensão adicional na exibição dos gráficos.
  - Muito mais efetivo para *display* do que para impressão.
  - Devem ser interativos (*pan*, *zoom*, rotação, etc.)

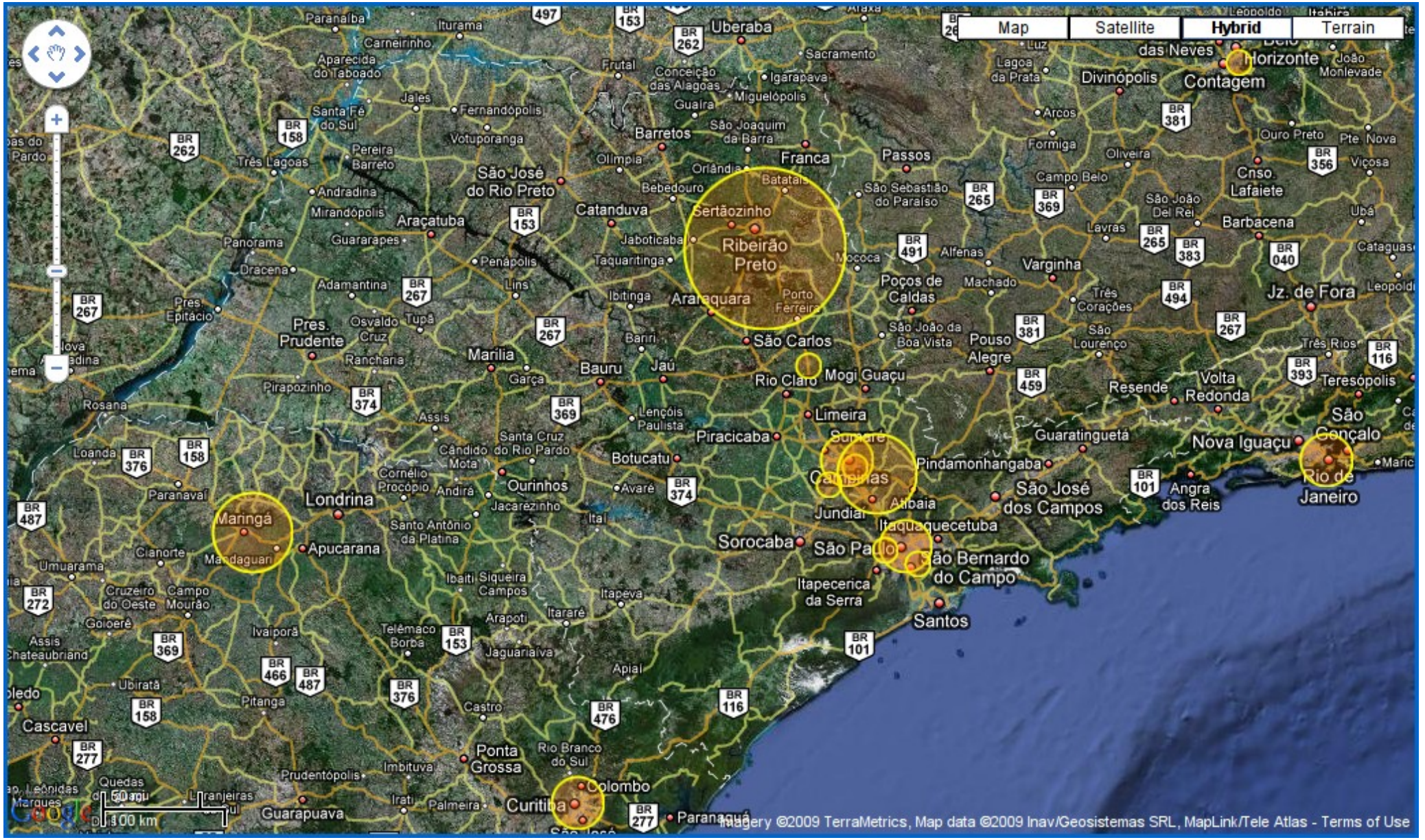
- **Mapas**
- Idéia básica: plotagem de elementos sobre coordenadas geográficas.
  - Valores, categorias, etc. podem ser representados como ícones, pixels, etc.
  - Devem ser interativos (*pan*, *zoom*).

Técnicas de Visualização: Mapas




```
42.1695 -87.9588 12
37.0 127.5 12
-21.3167 -47.75 12
39.9289 116.3883 6
41.0186 28.9647 6
-23.0 -46.85 6
```

Técnicas de Visualização: Mapas



Técnicas de Visualização: Comentários

- Esta taxonomia é incompleta e imperfeita.
- Técnicas podem pertencer a mais de uma categoria ou mesmo usar elementos de várias.
- Implementação de técnicas deve considerar também:
  - Interatividade com o gráfico em si (seleção, *drill-down*).
  - Interatividade com os dados usados para o gráfico (filtros, *queries*).

Extração e Pré-processamento de Dados de Segurança para Visualização

Motivação

- Atividade fundamental para segurança de sistemas: **MONITORAÇÃO!**
- Feita em passos:
  - Captura de tráfego e registro de eventos;
  - Armazenamento adequado;
  - Tratamento dos dados;
  - Análise e correlação de eventos.
- Informações úteis:
  - Falha de dispositivos, processos;
  - Alertas (disco cheio, *daemon* que não iniciou);
  - Ocorrência de ataques

O que é um "evento"?

- A interação resultante das atividades realizadas nos sistemas em geral (redes, dispositivos, computadores, programas) é denominada de **evento**.

- *"Um evento é uma modificação ou situação observável ocorrida em um ambiente por um período de tempo determinado."*

- *"Um evento pode ser um estado específico ou uma mudança de estado de um sistema."*

Raffael Marty. *Applied Security Visualization*. Addison Wesley, 2008.

Registros de eventos

- O registro de eventos em um sistema é chamado de *log.*
  - Coleta de dados de vários tipos de eventos de fontes diversas.

- Um *log* pode ser compreendido como um registro de transação ou auditoria que consiste de um ou mais arquivos do tipo texto ou em formatos específicos gerados por certas aplicações, que permite a visualização dos eventos ocorridos em um sistema.

- Mantêm o histórico de atividades, permitindo uma recuperação dos eventos ocorridos.

Fontes de dados: *Logs*

- *Logs* são capazes de representar, basicamente:
  - Atividades normais;
  - Alertas;
  - Erros.

- Devem prover informação suficiente para que o evento possa ser identificado e compreendido.

- Variam conforme a fonte geradora, mas devem listar ao menos informações de data/hora do evento, quem originou o evento e qual foi este evento.

Fontes geradoras de *logs*

- *Logs* não são limitados a sistemas específicos. Todo dispositivo ou programa deve ser capaz de gerar mensagens indicativas sobre suas operações.

- Alguns sistemas/dispositivos capazes de gerar *logs*:
  - Dispositivos de rede;
  - Sistema operacional;
  - Aplicações;
  - *Firewall*;
  - *IDS;*
  - *Antivirus.*

Logs: Tráfego de Rede

- Informações que identificam conexões entre máquinas, permitem a reconstrução de tráfego suspeito ⟶ *dump*.

```
$ sudo tcpdump -nXei en1 port 80

16:51:22.432458 00:21:29:d7:9a:ea > 00:19:e3:d3:d0:83, ethertype IPv4
(0x0800), length 74: 69.147.83.33.80 > 192.168.1.100.55134: S
3338542689:3338542689(0) ack 1256781315 win 65535 <mss 1460,nop,wscale
3,sackOK,timestamp 199158025 469452092>

[…] //Payload de outro pacote

0x0030:  0bde e909 4745 5420 2f20 4854 5450 2f31  ....GET./.HTTP/1
0x0040:  2e31 0d0a 486f 7374 3a20 7777 772e 6672  .1..Host:.www.fr
0x0050:  6565                                     ee
```

Logs: Tráfego de Rede (Exemplo)

- *Timestamp:* `16:51:22.432768`
- *MAC addresses:* `b1:fe::d3:b0:d3:00 > 00:21:29:d7:9a:ea`
- *Network protocol:* `ethertype IPv4 (0x0800)`
- *Packet length:* `length 521`
- *IP, port src > IP, port dst:* `192.168.1.100.55134 > 69.147.83.33.80`
- *Flags:* `S, ack`
- *Sequence number:* `3338542689`
- *ACK number:* `1256781315`
- *Window size:* `win 65535`
- *MSS:* `mss 1460`
- *Payload:* `GET / HTTP/1.1 Host: www.freebsd.org`

Logs: Sistema Operacional

- Mensagens informativas sobre S.O., hardware, aplicações.

```
Oct  1 15:04:40 Macintosh kernel[0]: sleep

Oct  1 13:16:14 Macintosh kernel[0]: AirPort: Link
Down on en1

Oct  1 16:40:52 Macintosh sudo[27285]:    andre :
TTY=ttys006 ; PWD=/Users/andre ; USER=root ;
COMMAND=/usr/sbin/tcpdump -ni en1

Oct  1 10:21:33 note hp_LaserJet_4250[26584]:
CMSCreateDataProviderOrGetInfo : Malformed
colorspace
```

Logs: Sistema Operacional (Exemplo)

Exemplo de *log unix-like (syslog):*

- *Timestamp:* `Oct  1 13:16:14`
- *Hostname:* `Macintosh`
- *Log generator agent:* `kernel`
- *Trigger service/device:* `AirPort`
- *Message:* `Link Down on en1`

Logs: Aplicações

- Registros acerca do funcionamento de uma aplicação (mensagens de início e finalização de um serviço, acesso a recursos, erros). Pode usar o mecanismo de *logging* do S.O.

```
Oct  5 05:21:33 asgard sshd[2393]: Server listening on
0.0.0.0 port 22.

Oct  1 15:11:01 Macintosh quicklook[26855]: unzip:
cannot find zipfile directory in one of
/Papers/SBSeg2009/SBSEG-App.odp

Oct  5 07:27:12 asgard sshd[3210]: Invalid user apple
from X.Y.Z.132
Oct  5 07:27:12 asgard sshd[3210]: Failed password for
invalid user apple from X.Y.Z.132 port 39427 ssh2
```

Logs: Firewall

- *Packet Filter (OpenBSD)*

```
20:54:01.006570 rule 0/(match) pass in on pcn0: 4.2.2.2 > 10.0.2.15: icmp: echo
reply
  0000: 4500 0054 0005 0000 fe01 aa91 0402 0202  E..T....♦.ª.....
  0010: 0a00 020f 0000 b667 2f3d 0000 4ac7 e328  ......¶g/=..JÇa(
  0020: 0000 0168 0809 0a0b 0c0d 0e0f 1011 1213  ...h............
  0030: 1415 1617                                ....

20:59:26.328625 rule 0/(match) pass in on lo0: ::1.36009 > ::1.22: [|tcp] [flowl
abel 0xdf2a8]
  0000: 600d f2a8 002c 0640 0000 0000 0000 0000  `.ò♦.,.@........
  0010: 0000 0000 0000 0001 0000 0000 0000 0000  ................
  0020: 0000 0000 0000 0001 8ca9 0016 c06d b503  .........c..Amµ.
  0030: 0000 0000                                ....

20:59:47.329984 rule 0/(match) pass in on lo0: 10.0.2.15.44199 > 10.0.2.15.22: S
 3644336743:3644336743(0) win 16384 <mss 33164,nop,nop,sackOK,nop,wscale 0,[|tcp
]> (DF) [tos 0x10]
  0000: 4510 0040 3d4b 4000 4006 e53f 0a00 020f  E..@=K@.@.å?....
  0010: 0a00 020f aca7 0016 d938 2a67 0000 0000  ....¬§..U8*g....
  0020: b002 4000 b0cc 0000 0204 818c 0101 0402  °.@.°I..........
  0030: 0103 0300                                ....
```

Logs: IDS

- Snort:

```
[**] [1:483:5] ICMP PING CyberKit 2.2 Windows [**]
[Classification: Misc activity] [Priority: 3]
08/01-03:37:36.117652 10.10.10.2 -> 192.168.20.29
ICMP TTL:121 TOS:0x0 ID:49936 IpLen:20 DgmLen:92
Type:8  Code:0  ID:16009   Seq:3967  ECHO
[Xref => http://www.whitehats.com/info/IDS154]
```

Logs: Antivirus

- ClamAV

```
Fri Oct  2 00:12:34 2009 -> ------------------------------------
Fri Oct  2 00:12:34 2009 -> Current working dir is
/opt/local/share/clamav
Fri Oct  2 00:12:34 2009 -> Max retries == 3
Fri Oct  2 00:12:34 2009 -> ClamAV update process started at Fri Oct
2 00:12:34 2009
Fri Oct  2 00:12:34 2009 -> Using IPv6 aware code
Fri Oct  2 00:12:35 2009 -> Querying current.cvd.clamav.net
Fri Oct  2 00:12:35 2009 -> TTL: 447
Fri Oct  2 00:12:35 2009 -> Software version from DNS: 0.95.2
Fri Oct  2 00:12:35 2009 -> main.cvd version from DNS: 51
Fri Oct  2 00:12:35 2009 -> main.cvd is up to date (version: 51, sigs:
545035, f-level: 42, builder: sven)
Fri Oct  2 00:12:35 2009 -> daily.cvd version from DNS: 9861
Fri Oct  2 00:12:35 2009 -> daily.cld is up to date (version: 9861,
sigs: 81209, f-level: 43, builder: guitar)
```

Logs: Antivirus 2

```
00.exe: Trojan.Dropper-20825 FOUND
2.exe: Trojan.Crypt-215 FOUND
Book_July_052009.JPG_.exe: Trojan.Packed-92 FOUND
a3.exe: Trojan.Agent-121026 FOUND
aa1.exe: Worm.Mytob-73 FOUND
amor.exe: Trojan.Downloader.Banload-5824 FOUND
wd.exe: Trojan.ShellHook-2 FOUND

----------- SCAN SUMMARY -----------
Known viruses: 1170581
Engine version: 0.95.2
Scanned directories: 0
Scanned files: 44
Infected files: 29
Data scanned: 13.17 MB
Data read: 11.35 MB (ratio 1.16:1)
Time: 14.856 sec (0 m 14 s)
```

Logs: Segurança

- Outros tipos de *log* voltados à descoberta de eventos de segurança (aplicações específicas).

- Objetivo ⟶ Monitorar:
  - Acessos indevidos a sistemas e redes;
  - Disseminação de *malware*;
  - Tráfego de ataque.

- Tecnologias:
  - *Honeypots;*
  - *Fluxos.*

Logs: Segurança (*Honeyd*)

- *Honeypot* de baixa interatividade → análise de tendências.

```
Connection established: tcp (10.0.0.1:6324 -
192.168.1.1:23) <-> scripts/router-telnet.pl
E(10.0.0.1:6324 - 192.168.1.1:23): Attempted
login: root/root123
```

- Estado da conexão
- Protocolo
- Endereços IP e portas comunicantes
- Emulador utilizado
- Mensagem resultante da sessão

Logs: Segurança (*Nepenthes*)

- *Honeypot* de baixa interatividade → coleta de *malware*

```
[12042009 16:36:51 warn module] Unknown NETDDE exploit 76 bytes
State 1
[12042009 16:36:51 warn module] Unknown SMBName exploit 0 bytes
State 1
[12042009 16:36:51 info handler dia] Unknown DCOM request,
dropping
[12042009 16:36:57 info sc handler] i = 1 map_items 2 , map =
port
[12042009 16:36:57 info sc handler] bindfiletransfer::amberg ->
9988
[12042009 16:36:57 info sc handler] bindfiletransfer::amberg ->
w.x.y.z:9988
[12042009 16:36:57 info down mgr] Handler creceive download
handler will download creceive://w.x.y.z:9988/0
[12042009 16:37:12 info mgr submit] File
9604e9c99768c5cd2deb108935356196 has type MS-DOS executable PE
for MS Windows (GUI) Intel 80386 32-bit
```

Logs: Segurança (*Flows*)

- *NetFlow:* ferramenta de coleta e medição de tráfego de roteadores e *switches*.
- Fluxo: Sequência unidirecional de pacotes entre um dado par de origem-destino ambos definidos pelo IP e portas.

srcaddr	dstaddr	srcport	dstport	Pkts	Bytes	first	last	flags	prot
X.Y.255.238	X.Y.0.193	19752	20894	453	90600	2009-10-02 09:26:56	2009-10-02 09:27:16	16	17
X.23.20.17	X.Y.205.190	80	2053	14	13983	2009-10-02 09:54:59	2009-10-02 09:54:59	27	6
X.Y.205.232	A.B.163.99	2506	80	4	906	2009-10-02 09:49:28	2009-10-02 09:49:28	26	6
X.Y.201.11	X.Y.205.232	143	2415	8	795	2009-10-02 09:43:22	2009-10-02 09:43:22	27	6
X.Y.2.107	X.Y.207.102	4158	1198	3	128	2009-10-02 09:27:55	2009-10-02 09:27:55	22	6
X.Y.208.14	X.154.56.76	2165	80	10	5502	2009-10-02 09:31:32	2009-10-02 09:31:42	26	6
X.Y.208.36	X.Y.121.120	443	50828	8	2586	2009-10-02 09:32:20	2009-10-02 09:32:20	27	6
B.C.13.104	X.Y.208.14	80	2613	1	40	2009-10-02 09:53:56	2009-10-02 09:53:56	17	6
X.Y.208.14	A.B.13.110	2777	80	5	900	2009-10-02 09:53:56	2009-10-02 09:53:56	27	6
Z.W.163.149	X.Y.201.91	80	46908	4	859	2009-10-02 09:36:29	2009-10-02 09:36:30	27	6
X.W.210.59	X.Y.216.129	80	35202	2	522	2009-10-02 09:46:51	2009-10-02 09:46:57	25	6
X.Y.208.37	X.Y.201.1	35797	53	1	54	2009-10-02 09:33:34	2009-10-02 09:33:34	16	17
D.E.223.64	X.Y.201.1	53	65310	1	311	2009-10-02 09:52:47	2009-10-02 09:52:47	16	17

Problemas com integridade das informações

- Muitos tipos de dados diferentes e que servem para monitoração e administração de sistemas.

- Grande quantidade de dados → tratamento para extração de informações úteis para segurança.

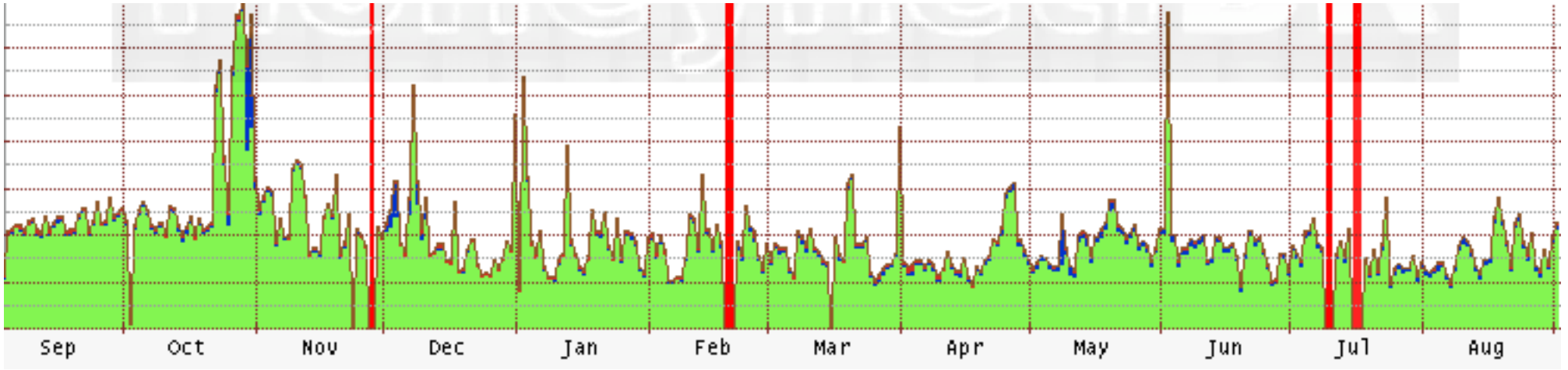
- Problemas:
  - Geração dos *logs* (má configuração);
  - Armazenamento incorreto (falta de espaço, informações incompletas ou sobrescritas, arquivos corrompidos);

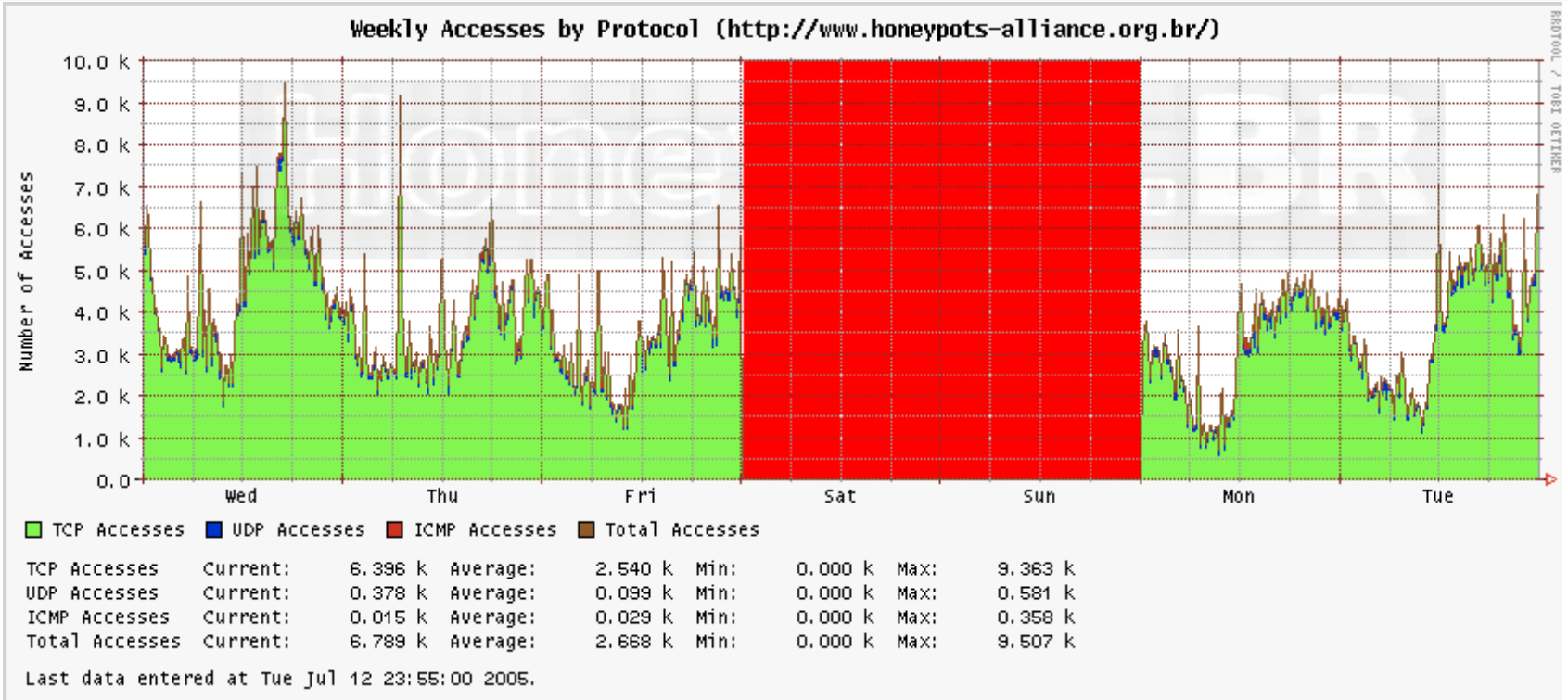
Integridade: Dados incompletos

- Podem causar interpretação errônea:
  - Falsos alarmes em eventos normais;
  - Nenhum dado a ser visualizado em caso de evento suspeito.
- Ocorrem quando os dados coletados sofrem parada abrupta no processo de coleta, ou erro no armazenamento.
- Prejudica completamente o resultado final.
- Impossibilita o correlacionamento de eventos.

- Solução: projeto adequado da arquitetura de coleta e armazenamento; monitoração dos processos envolvidos.

Integridade: Dados incompletos (Exemplo)





Integridade: Erros de sincronização

- Causados pela falta de sincronização entre os relógios dos componentes do sistema.
- Impossibilitam o correlacionamento.
- Inviabilizam a visualização correta dos eventos (*timeline*).
- Solução: configurar NTP!

Integridade: Erros de sincronização (Exemplo 1)


```
22:17:17.938078 IP 10.0.0.103.49308 > XXX.YYY.WWW.ZZZ.22: S 778826581:778826581(0) win
   65535
22:17:18.016194 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: S 3350360119:3350360119(0) ack
   778826582
22:17:18.016270 IP 10.0.0.103.49308 > XXX.YYY.WWW.ZZZ.22: . ack 1 win
22:17:18.095738 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: P 1:21(20) ack 1
(...)
22:17:23.181390 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: . ack 1213 win
22:17:24.967674 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: P 1725:1805(80) ack 1213
22:17:24.967745 IP 10.0.0.103.49308 > XXX.YYY.WWW.ZZZ.22: . ack 1805 win
22:17:27.443235 IP 10.0.0.103.49308 > XXX.YYY.WWW.ZZZ.22: P 1213:1357(144) ack 1805
22:17:27.521359 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: . ack 1357 win
22:17:29.227115 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: P 1805:1885(80) ack 1357
22:17:29.227184 IP 10.0.0.103.49308 > XXX.YYY.WWW.ZZZ.22: . ack 1885 win
22:17:30.995193 IP 10.0.0.103.49308 > XXX.YYY.WWW.ZZZ.22: P 1357:1501(144) ack 1885
22:17:31.077129 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: . ack 1501 win
22:17:33.335692 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: P 1885:1965(80) ack 1501
22:17:33.335762 IP 10.0.0.103.49308 > XXX.YYY.WWW.ZZZ.22: . ack 1965 win
22:17:33.337146 IP 10.0.0.103.49308 > XXX.YYY.WWW.ZZZ.22: F 1501:1501(0) ack 1965
22:17:33.415258 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: F 1965:1965(0) ack 1502
22:17:33.703246 IP XXX.YYY.WWW.ZZZ.22 > 10.0.0.103.49308: F 1965:1965(0) ack 1502
22:17:33.703320 IP 10.0.0.103.49308 > XXX.YYY.WWW.ZZZ.22: . ack 1966 win
```

Integridade: Erros de sincronização (Exemplo 2)


```
May 10 22:26:07 alvo sshd[16023]: Invalid user admin from aaa.bb.ccc.232
May 10 22:26:07 alvo sshd[16024]: input_userauth_request: invalid user admin
May 10 22:26:09 alvo sshd[16023]: pam_unix(sshd:auth): check pass; user unknown
May 10 22:26:09 alvo sshd[16023]: pam_unix(sshd:auth): authentication failure; logname=
   uid=0 euid=0 tty=ssh ruser= rhost=aaa.bb.ccc-232.dsl.telesp.net.br
May 10 22:26:09 alvo sshd[16023]: pam_succeed_if(sshd:auth): error retrieving information
   about user admin
May 10 22:26:12 alvo sshd[16023]: Failed password for invalid user admin from
   aaa.bb.ccc.232 port 60609 ssh2
May 10 22:26:14 alvo sshd[16023]: pam_unix(sshd:auth): check pass; user unknown
May 10 22:26:14 alvo sshd[16023]: pam_succeed_if(sshd:auth): error retrieving information
   about user admin
May 10 22:26:16 alvo sshd[16023]: Failed password for invalid user admin from
   aaa.bb.ccc.232 port 60609 ssh2
May 10 22:26:19 alvo sshd[16023]: pam_unix(sshd:auth): check pass; user unknown
May 10 22:26:19 alvo sshd[16023]: pam_succeed_if(sshd:auth): error retrieving information
   about user admin
May 10 22:26:21 alvo sshd[16023]: Failed password for invalid user admin from
   aaa.bb.ccc.232 port 60609 ssh2
May 10 22:26:21 alvo sshd[16024]: Connection closed by aaa.bb.ccc.232
May 10 22:26:21 alvo sshd[16023]: PAM 2 more authentication failures; logname= uid=0
   euid=0 tty=ssh ruser= rhost=aaa.bb.ccc-232.dsl.telesp.net.br
```


Integridade: Ruído nos *logs*

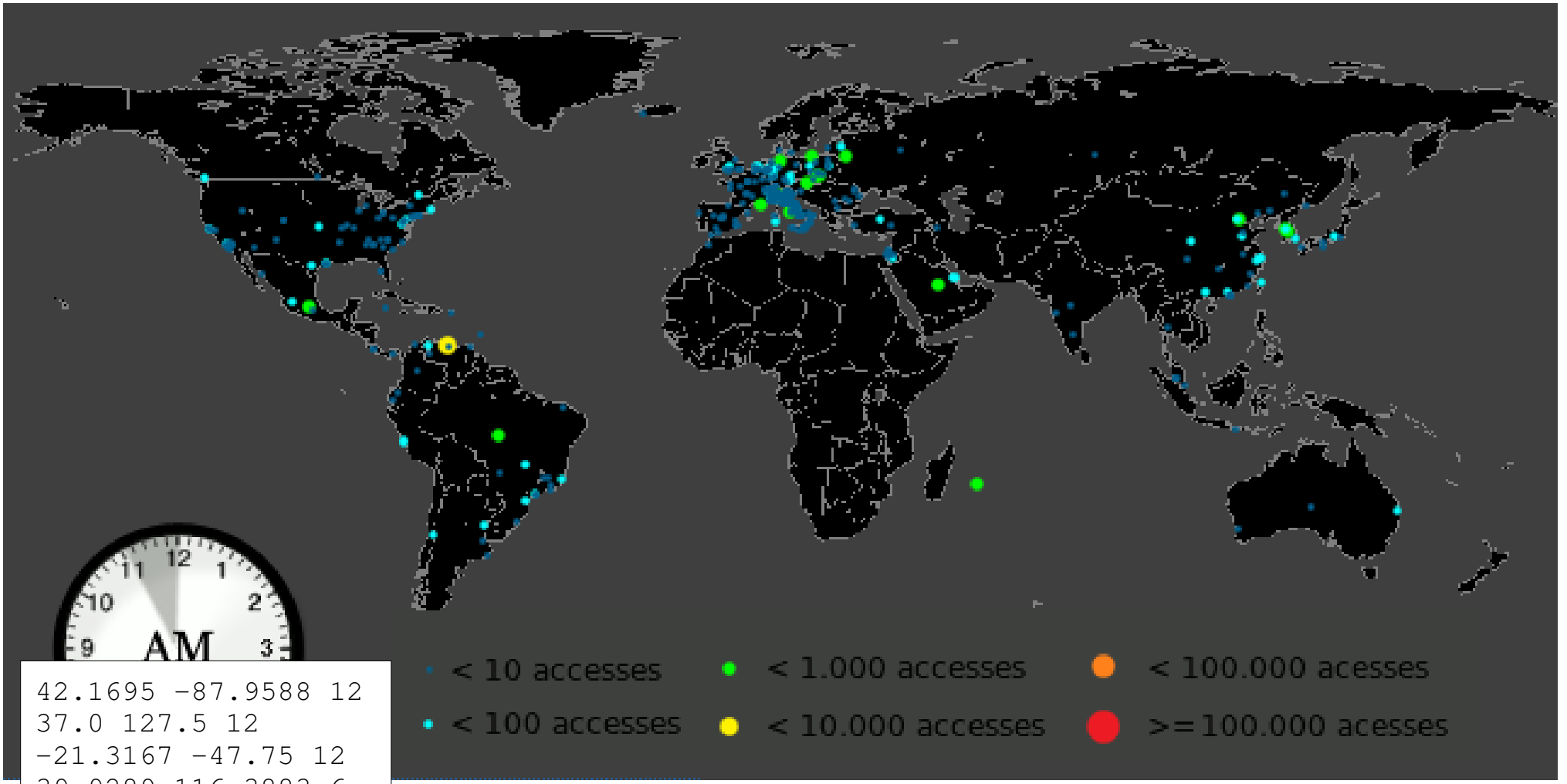
- Problema comum quando:
  - Filtros não são corretamente aplicados;
  - Administrador não utiliza o sistema de acordo.
- Eventos que não deveriam fazer parte dos *logs*, por não conter informações úteis para análise, aparecem!
- Atrapalham a visualização dos eventos importantes.
- Causam erros de interpretação.
- Ex.: Sistemas de análise de *malware* que registram chamadas de sistema das ferramentas utilizadas no processo.

Integridade: Ruído nos *logs* (Exemplo)

- 16:42:20.372103 IP 10.0.0.109.995 > 192.168.1.100.55088: P 4220351:4220354(3) ack 32 win 106 <nop,nop,timestamp 4269089901 469446676>
- 16:42:20.372135 IP 192.168.1.100.55088 > 10.0.0.109.995: . ack 4220354 win 65535 <nop,nop,timestamp 469446679 4269089901>
- 16:42:20.372456 IP 10.0.0.109.995 > 192.168.1.100.55088: P 4220354:4220884(530) ack 32 win 106 <nop,nop,timestamp 4269089901 469446676>
- 16:42:20.372480 IP 192.168.1.100.55088 > 10.0.0.109.995: . ack 4220884 win 65516 <nop,nop,timestamp 469446679 4269089901>
- 16:42:20.377495 IP 10.0.0.109.995 > 192.168.1.100.55088: P 4220884:4220908(24) ack 32 win 106 <nop,nop,timestamp 4269089902 469446676>
- 16:42:20.377530 IP 192.168.1.100.55088 > 10.0.0.109.995: . ack 4220908 win 65535 <nop,nop,timestamp 469446679 4269089902>
- **16:42:20.378172 IP 172.16.1.91.443 > 192.168.1.100.55113: . ack 117 win 46 <nop,nop,timestamp 1146678156 469446679>**
- **16:42:20.378282 IP 172.16.1.91.443 > 192.168.1.100.55113: P 1:155(154) ack 117 win 46 <nop,nop,timestamp 1146678156 469446679>**
- **16:42:20.378310 IP 192.168.1.100.55113 > 172.16.1.91.443: . ack 155 win 65535 <nop,nop,timestamp 469446679 1146678156>**
- 16:42:20.378407 IP 192.168.1.100.55088 > 10.0.0.109.995: P 32:63(31) ack 4220908 win 65535 <nop,nop,timestamp 469446679 4269089902>
- **16:42:20.379702 IP 192.168.1.100.55113 > 172.16.1.91.443: P 117:885(768) ack 155 win 65535 <nop,nop,timestamp 469446680 1146678156>16:42:20.425229 IP 172.16.1.91.443 > 192.168.1.100.55113: . 1603:3051(1448) ack 885 win 58 <nop,nop,timestamp 1146678227 469446680>**
- **16:42:20.429400 IP 172.16.1.91.443 > 192.168.1.100.55113: P 3051:4117(1066) ack 885 win 58 <nop,nop,timestamp 1146678227 469446680>**
- 16:42:20.530451 IP 10.0.0.109.995 > 192.168.1.100.55088: . ack 63 win 106 <nop,nop,timestamp 4269090093 469446679>
- 16:42:20.677989 IP 10.0.0.109.995 > 192.168.1.100.55088: . 4220908:4222326(1418) ack 63 win 106 <nop,nop,timestamp 4269090239 469446679>

Técnicas de Visualização com Aplicações a Dados de Segurança

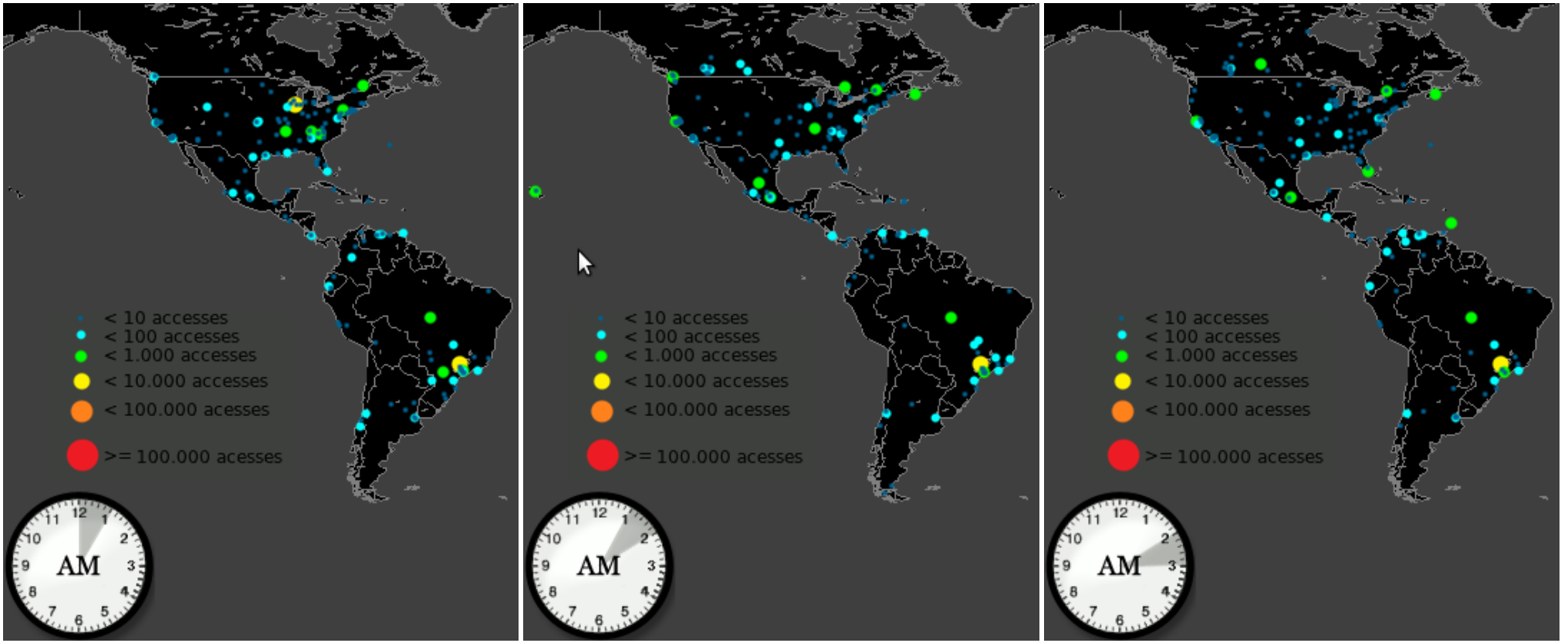
Dados Geográficos de Atividades Maliciosas




```
42.1695 -87.9588 12
37.0 127.5 12
-21.3167 -47.75 12
39.9289 116.3883 6
41.0186 28.9647 6
-23.0 -46.85 6
```

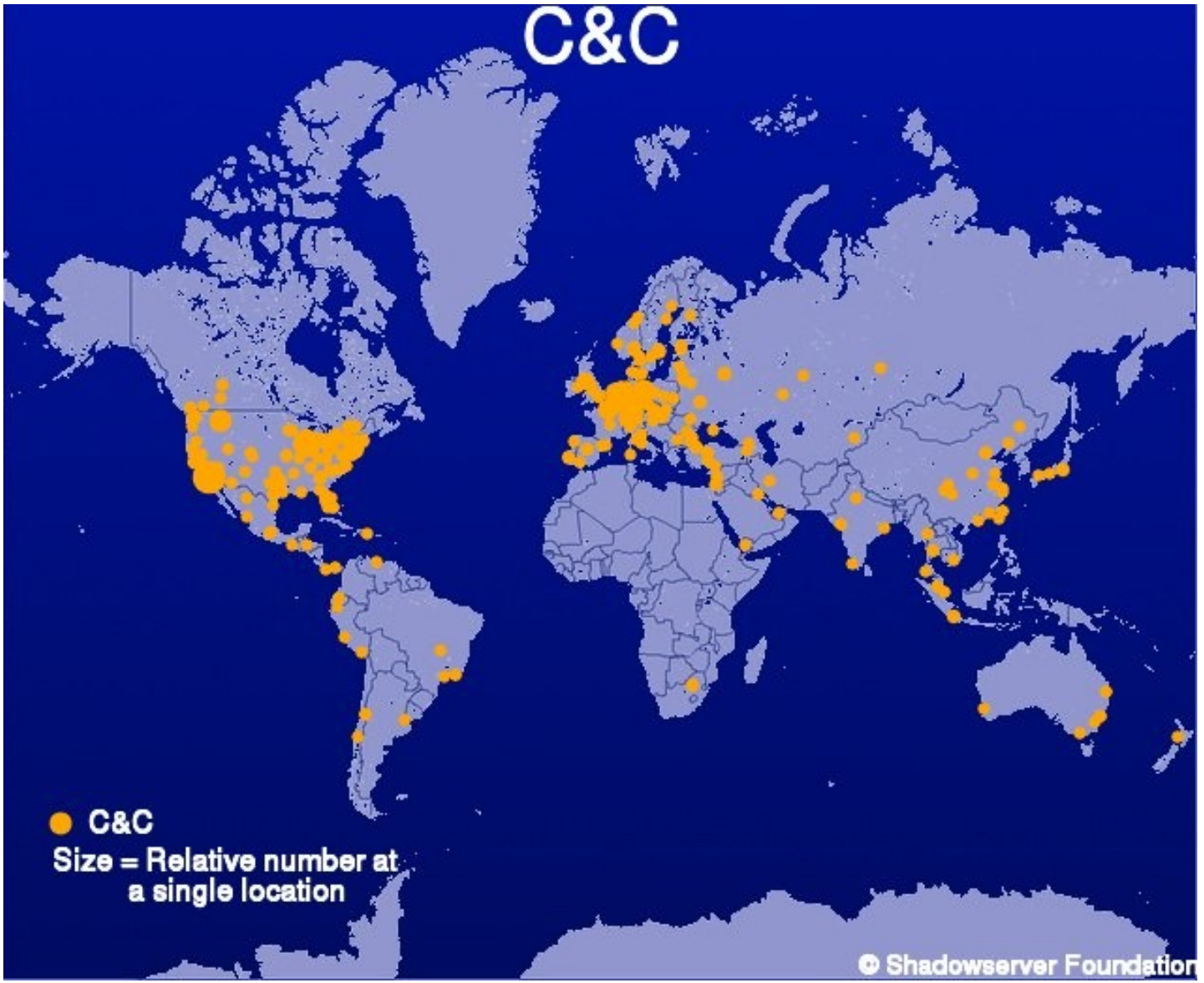
Estatísticas do Consórcio Brasileiro de Honeypots, http://www.dssi.cti.gov.br/dssi/statistics.html

Dados Geográficos de Atividades Maliciosas



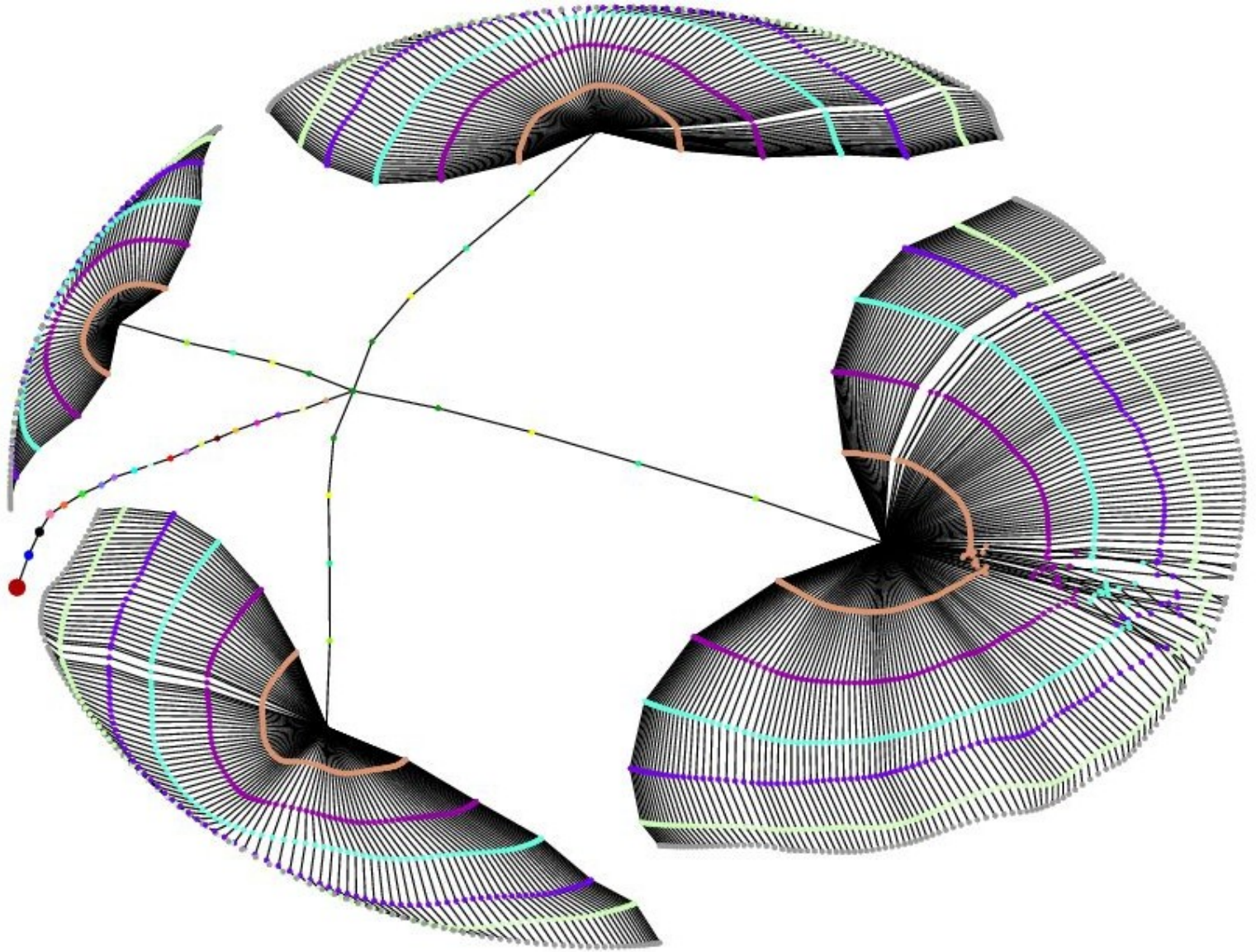
Estatísticas do Consórcio Brasileiro de Honeypots, http://www.dssi.cti.gov.br/dssi/statistics.html

Servidores de Controle de Botnets



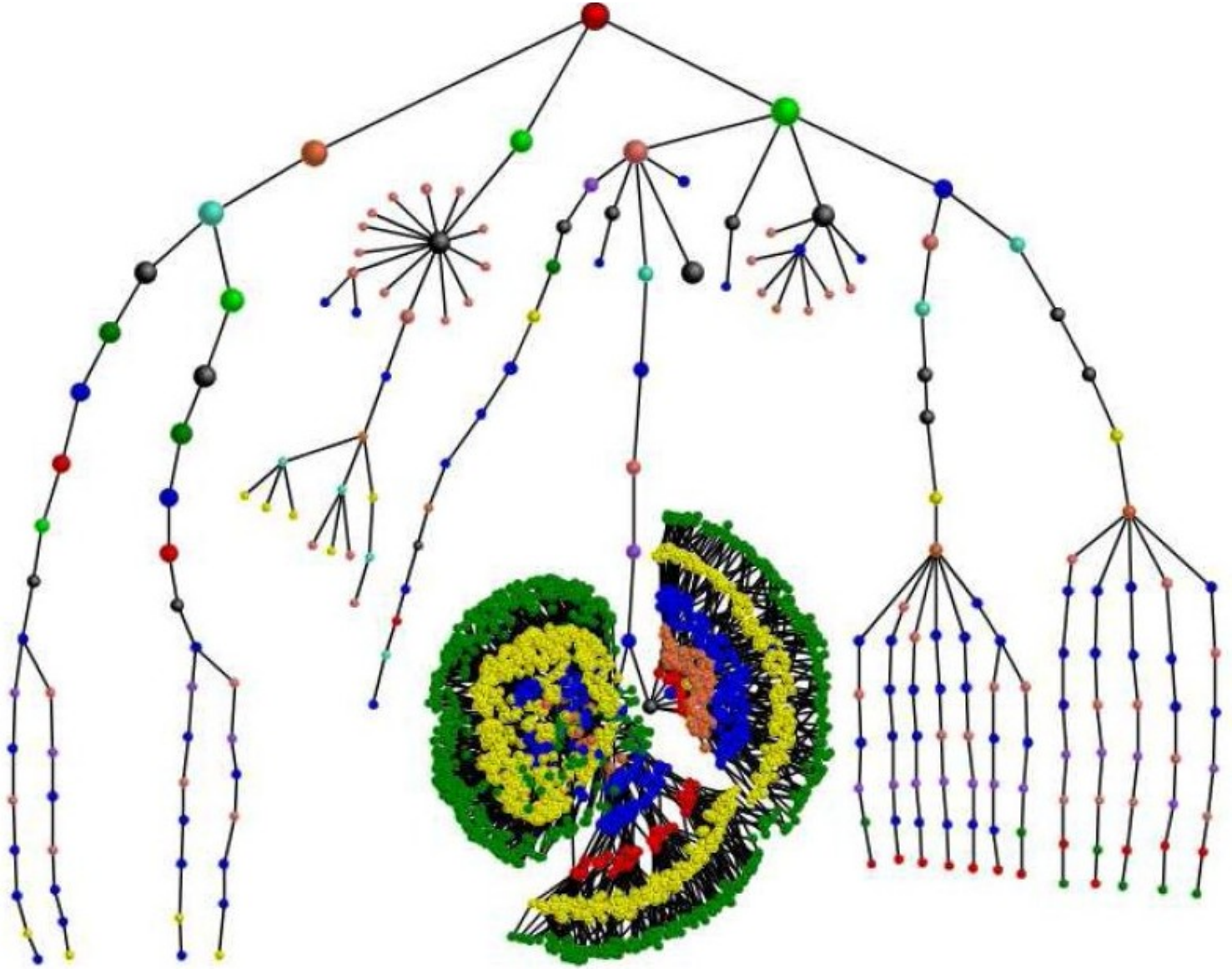
The Shadowserver Foundation, http://www.shadowserver.org

Identificação de Padrões em Campanhas de SPAM



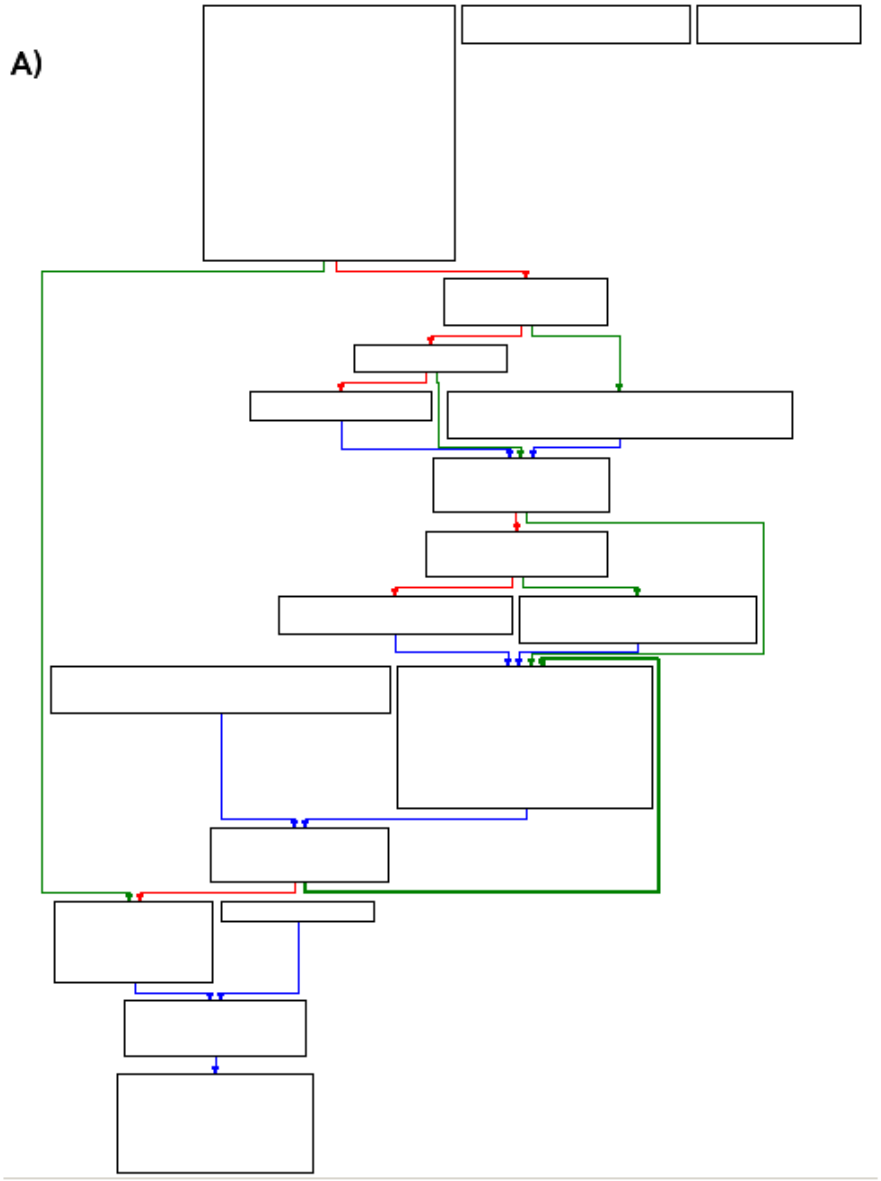
Projeto SpamPots, http://www.cert.br/docs/whitepapers/spampots (CERT.br e UFMG)

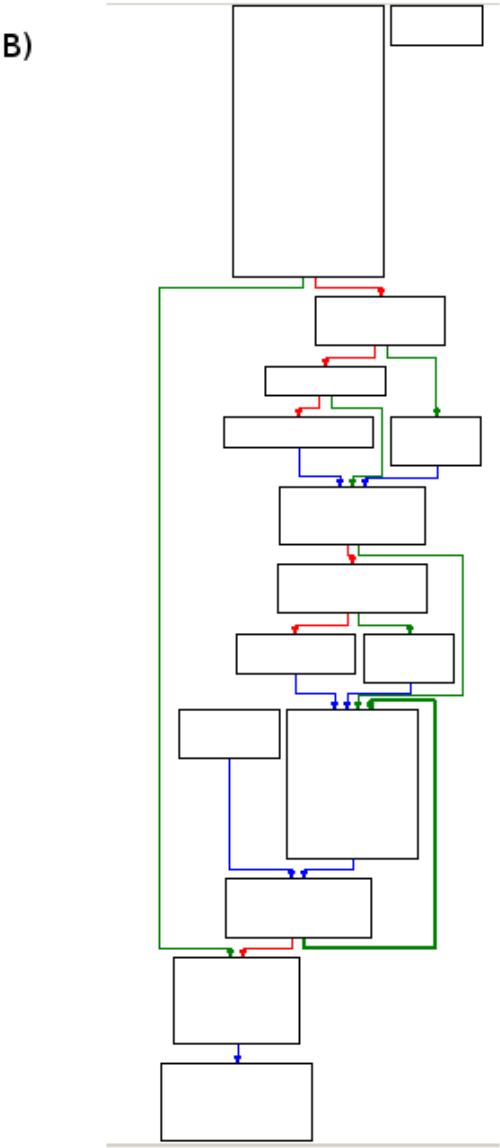
Identificação de Padrões em Campanhas de SPAM



Projeto SpamPots, http://www.cert.br/docs/whitepapers/spampots (CERT.br e UFMG)

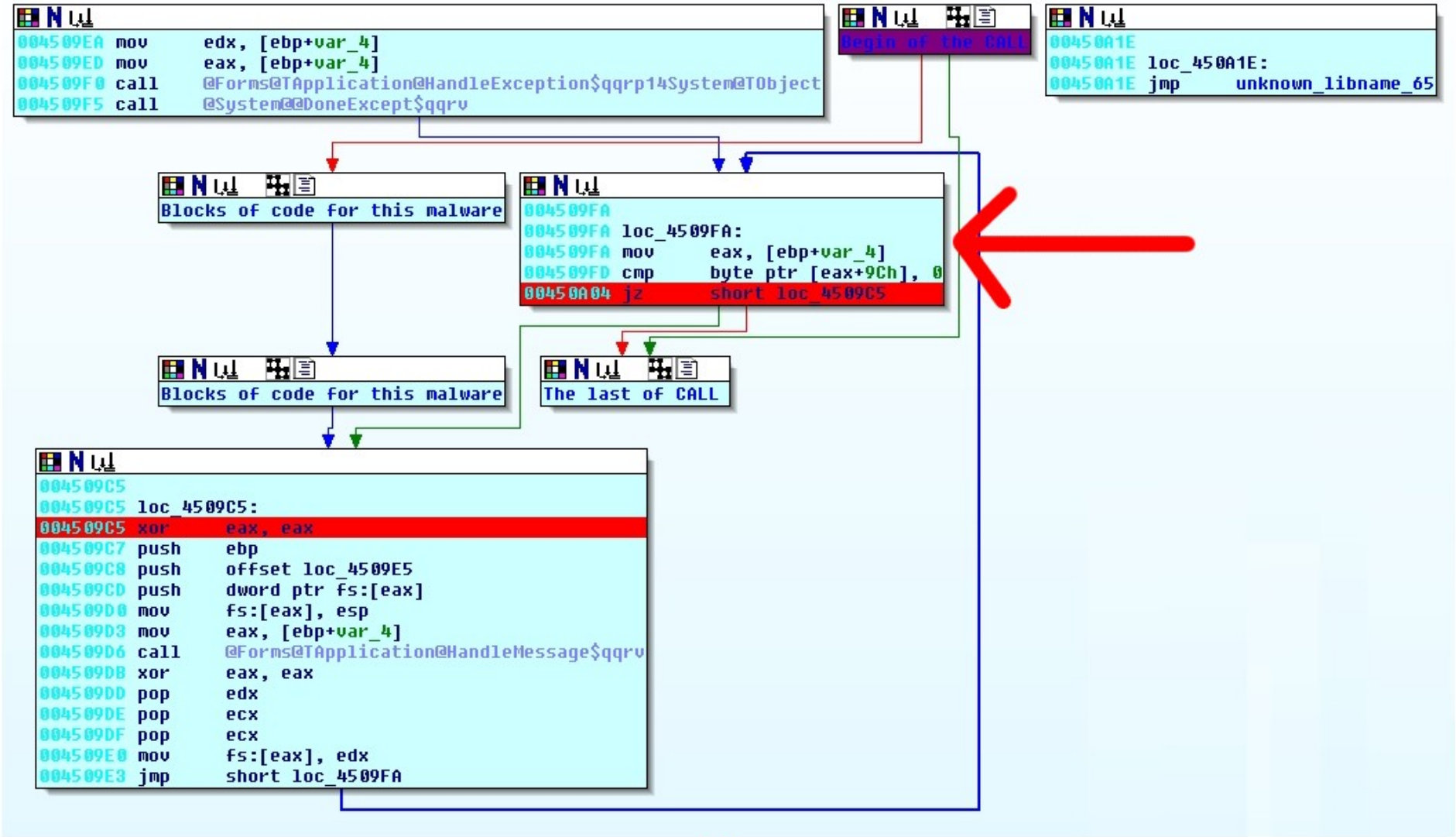
Partes do Fluxo de Execução de Malwares





Gráficos criados com IDA Pro (http://www.hex-rays.com/idapro/)

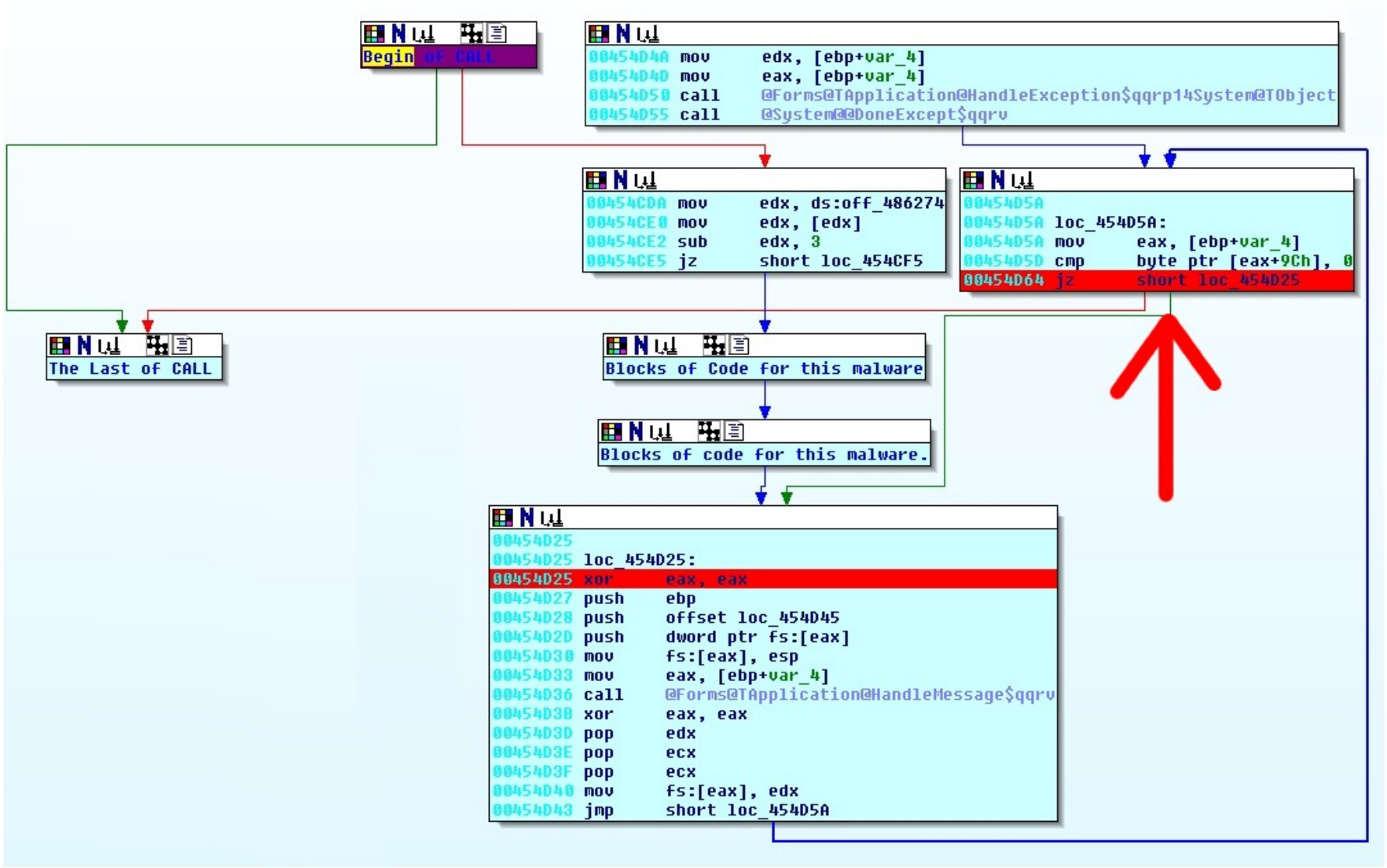
Partes do Fluxo de Execução de Malwares



(I)

Gráficos criados com IDA Pro (http://www.hex-rays.com/idapro/)

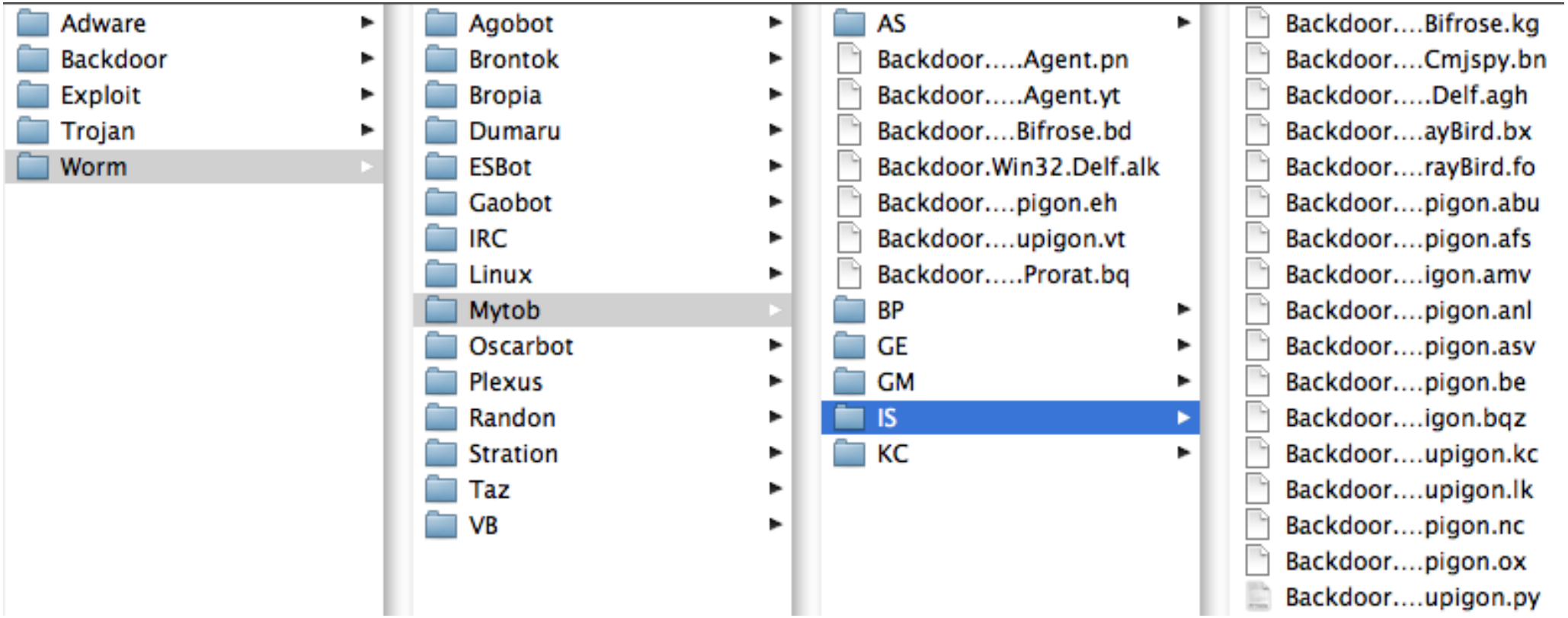
Partes do Fluxo de Execução de Malwares



(II)

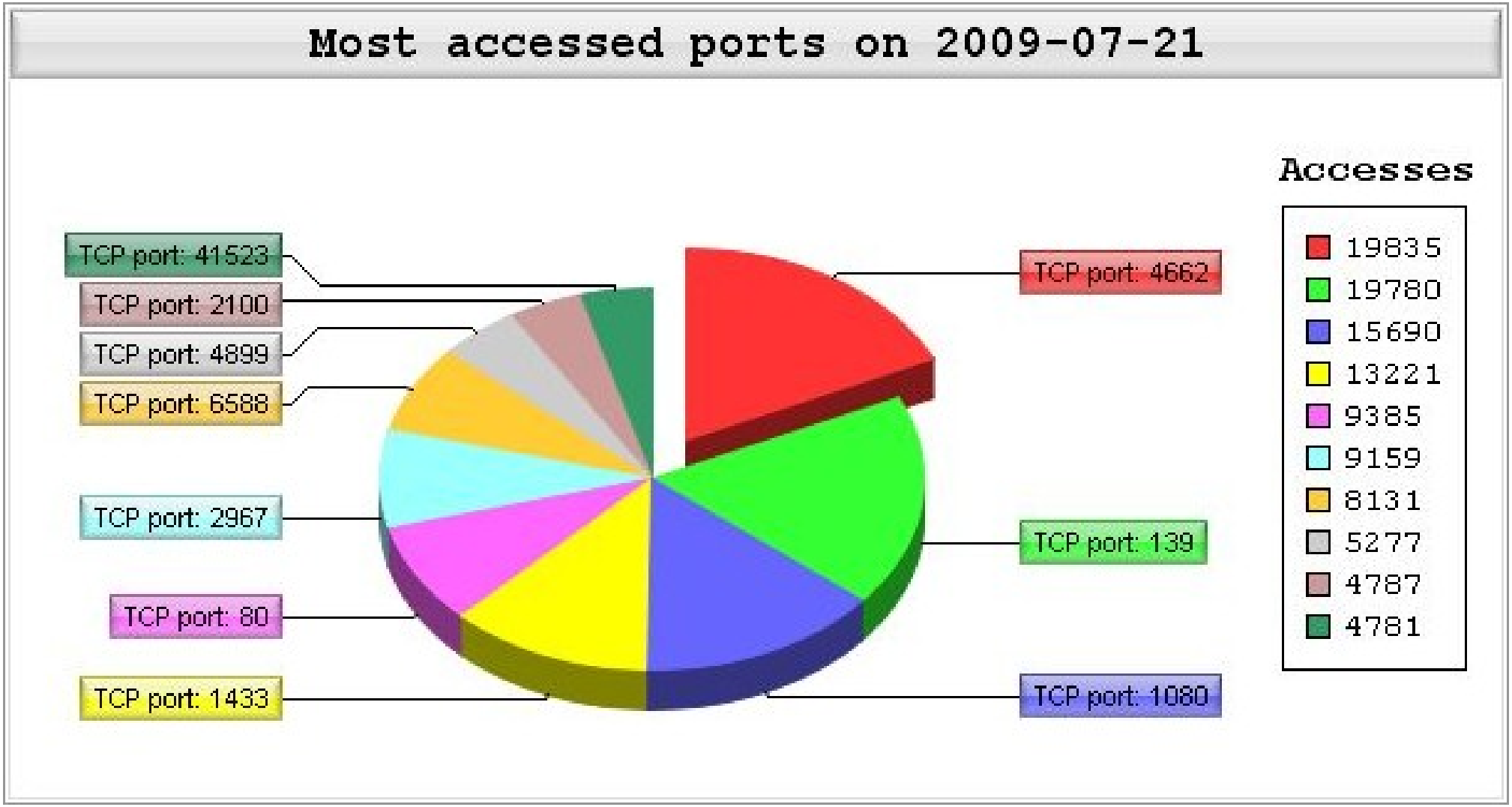
Gráficos criados com IDA Pro (http://www.hex-rays.com/idapro/)

Separação de Malware por Subclasses



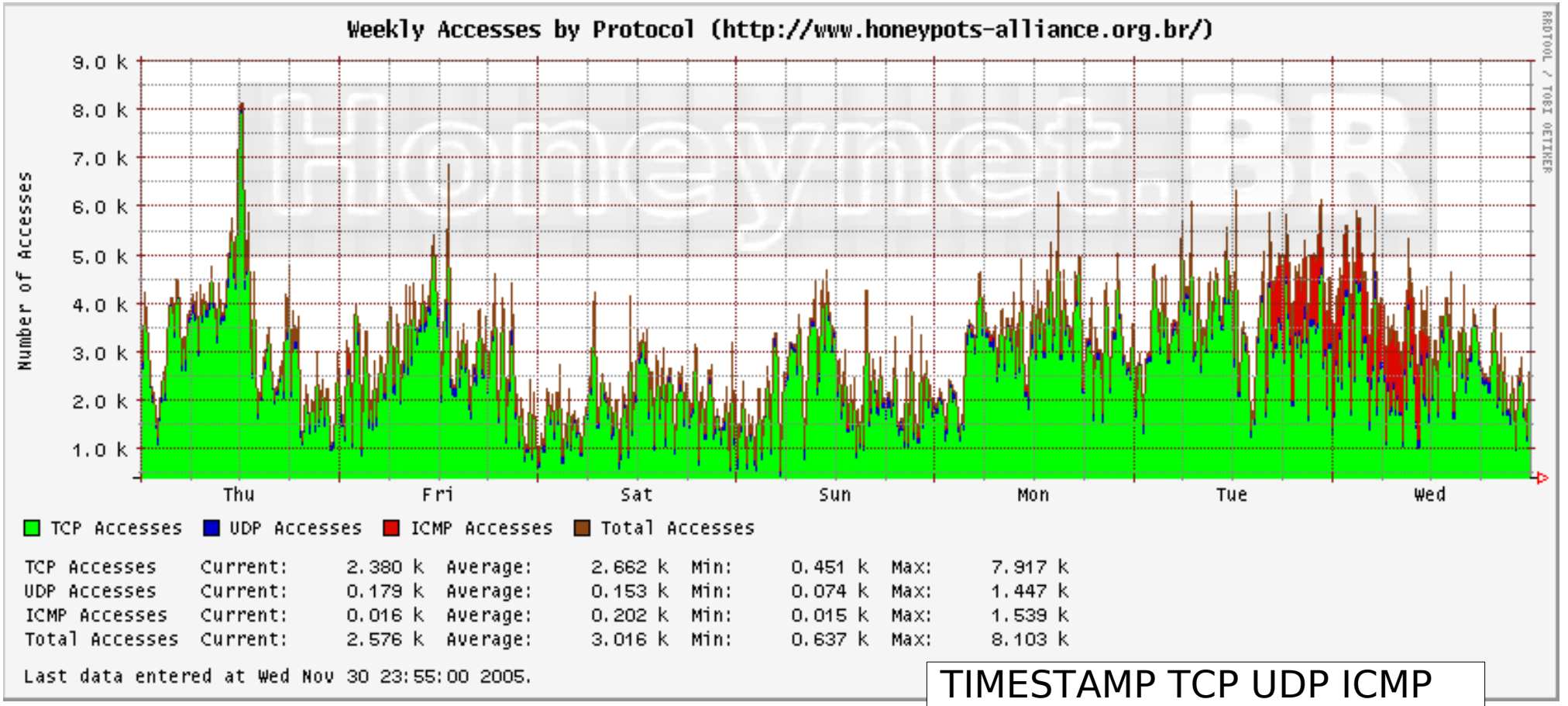
Script em Perl, identificadores gerados por ClamAV.

Portas acessadas em um Honeypot



Divisão de Segurança de Sistemas de Informação, CTI

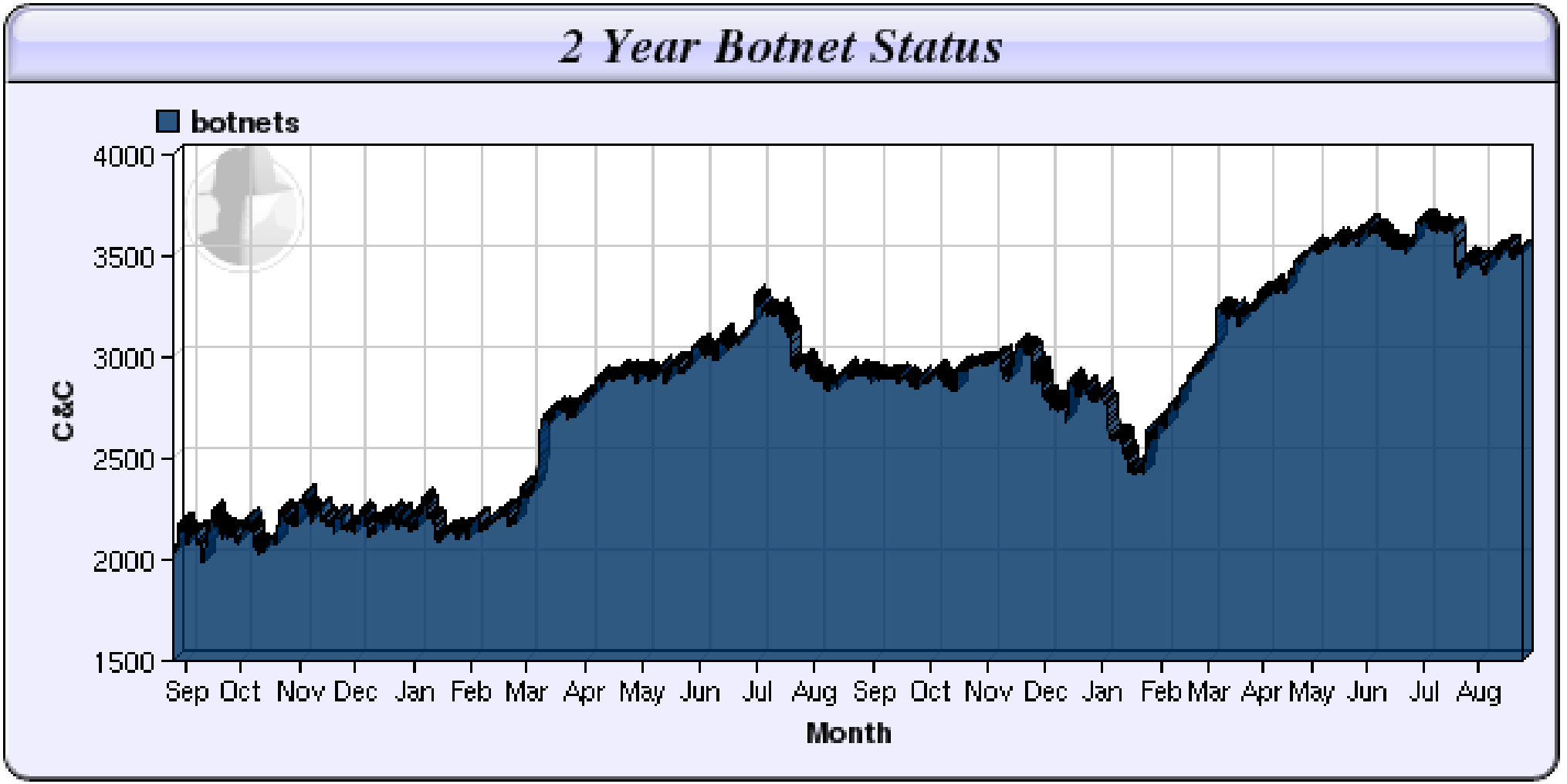
Quantidades de acessos a Honeypots



TIMESTAMP TCP UDP ICMP
1133308800 1247 156 1294
1133309100 1114 60 1285
1133309400 2760 125 1275
1133309700 1062 202 1266
1133310000 2753 117 1277

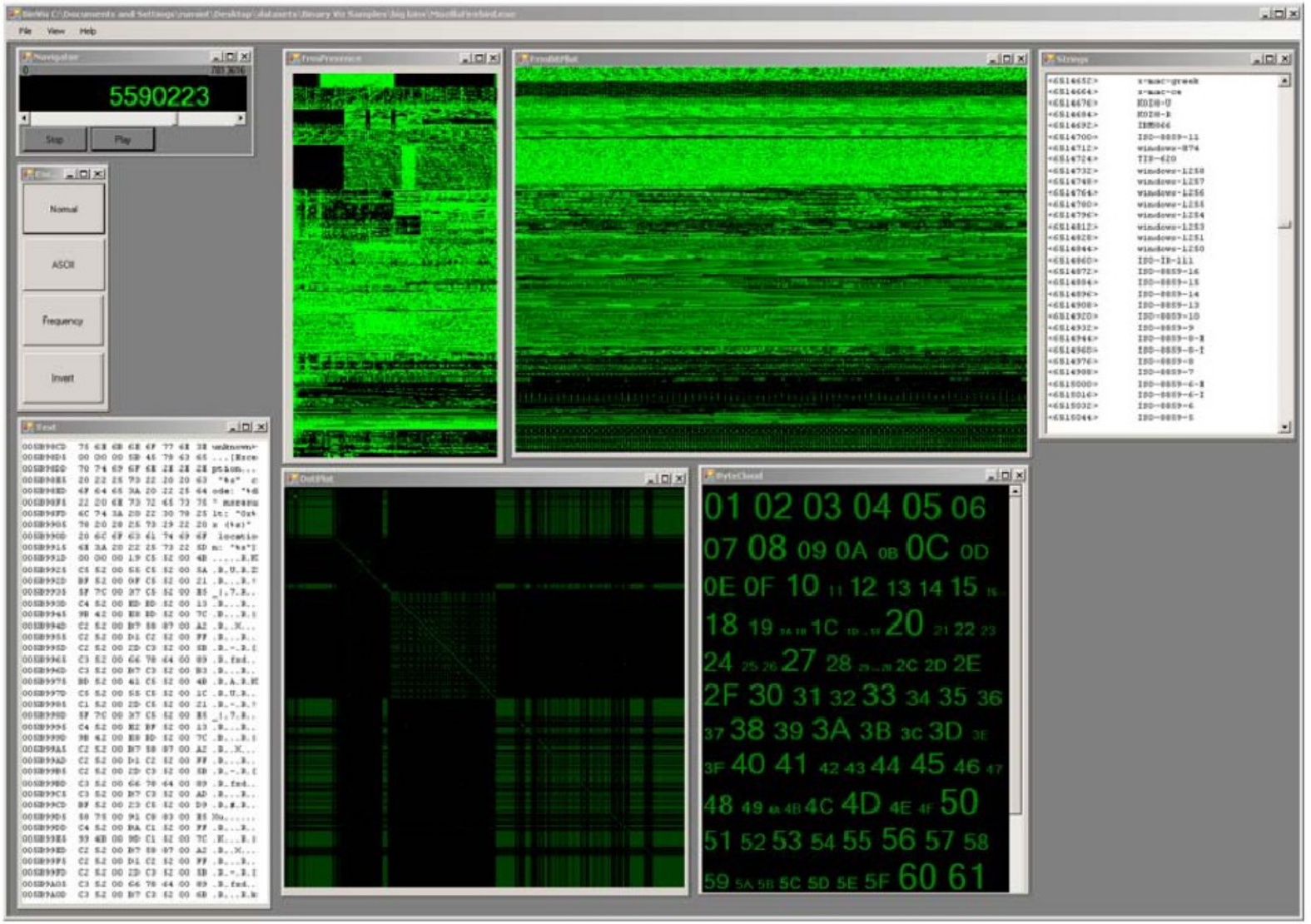
Divisão de Segurança de Sistemas de Informação, CTI

Servidores de Controle de Botnets Monitorados



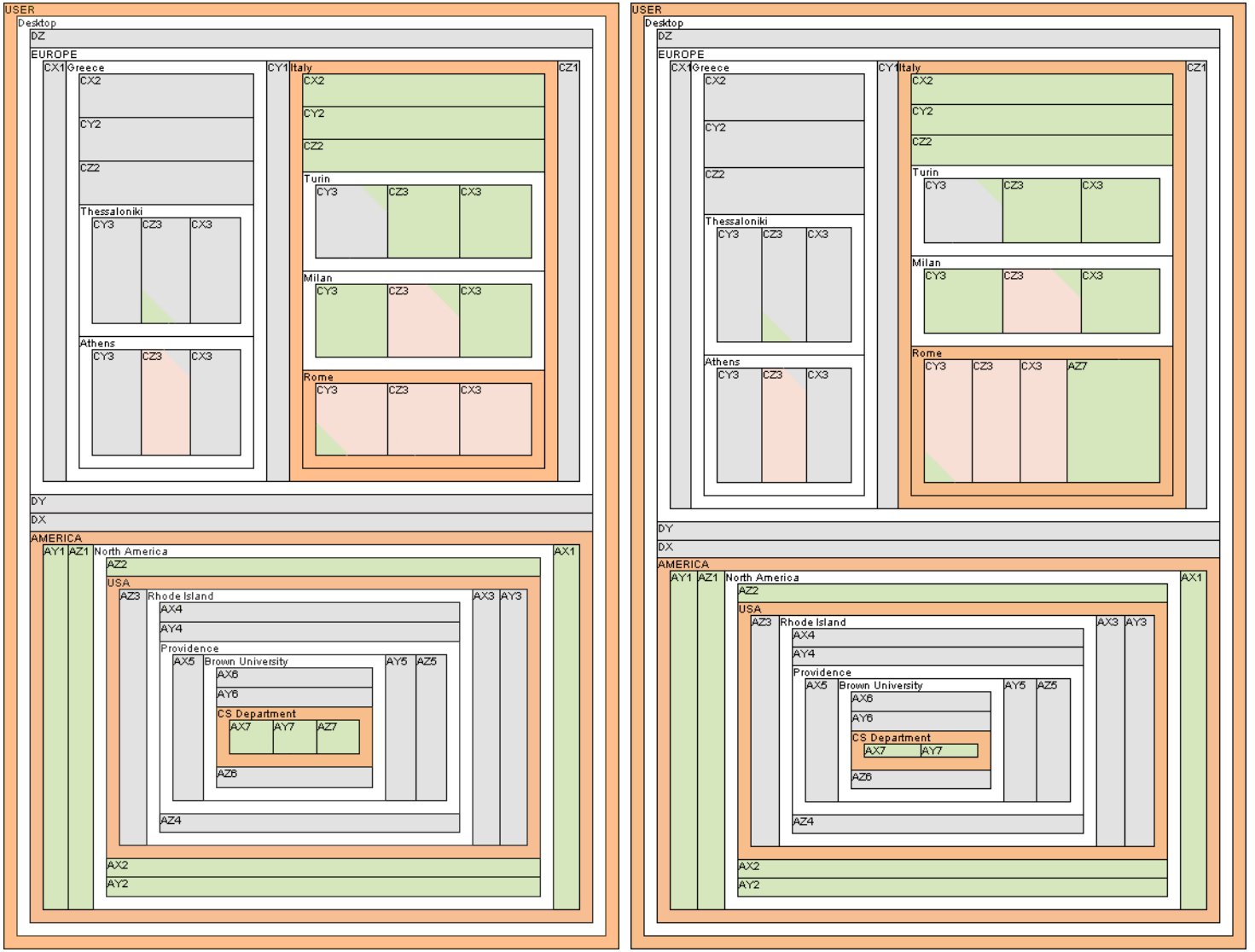
The Shadowserver Foundation, http://www.shadowserver.org

Conteúdo de Arquivos Binários



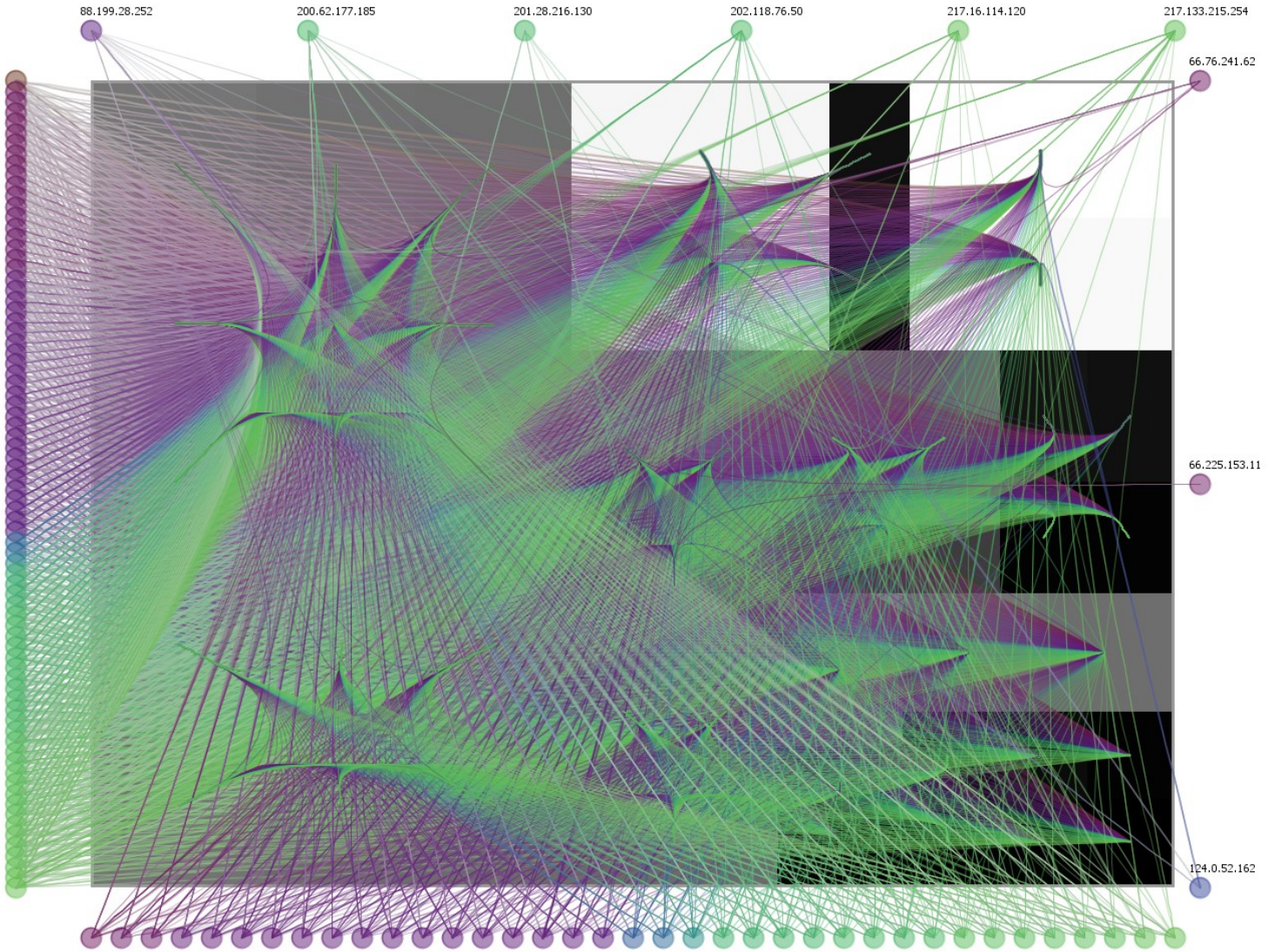
Gregory Conti, Erik Dean, Matthew Sinda, and Benjamin Sangster. *Visual Reverse Engineering of Binary and Data Files*. Visualization for Computer Security, VizSec 2008 (LNCS 5210)

Visualização de Controle de Permissões



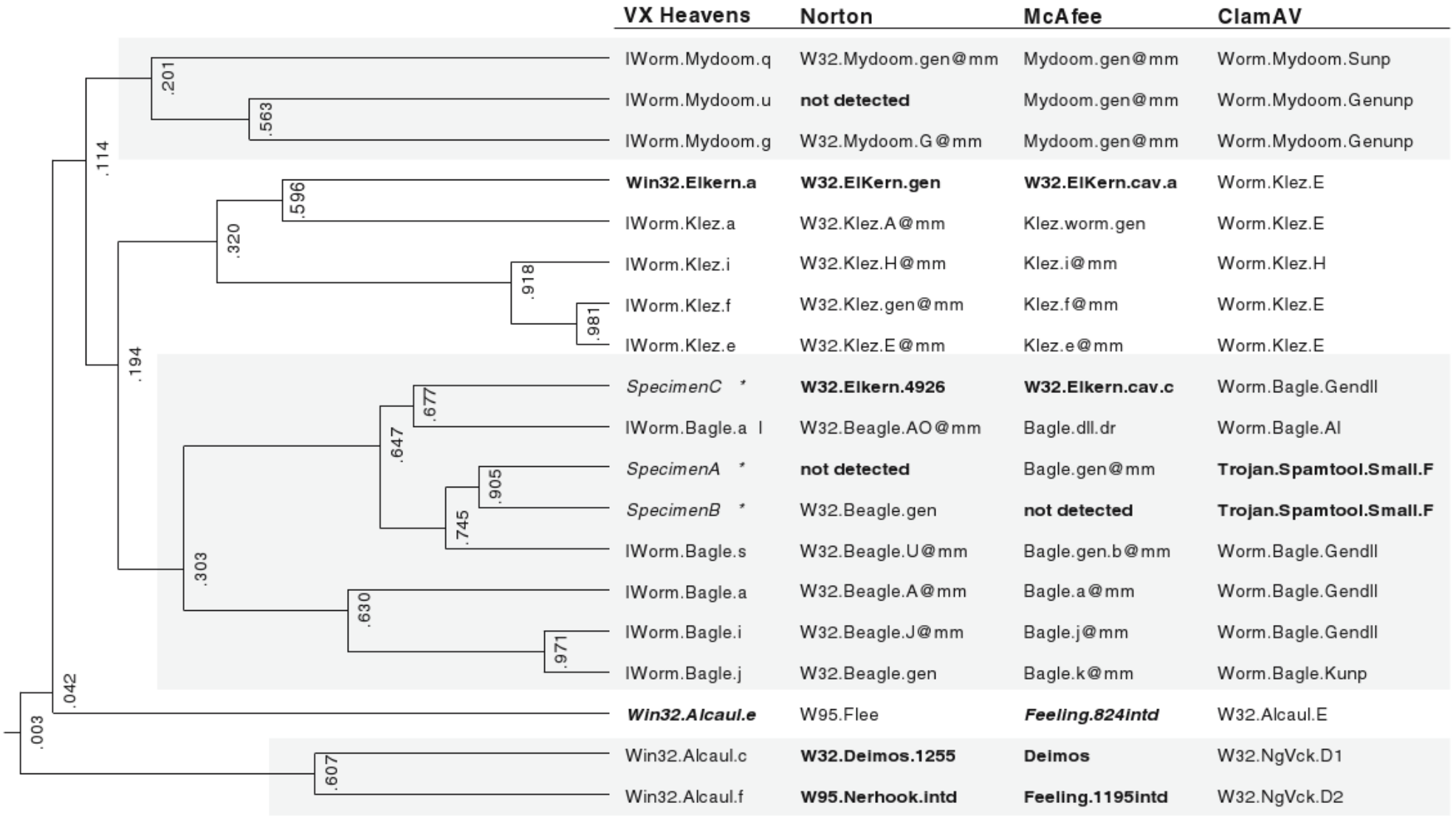
Alexander Heitzmann, Bernardo Palazzi, Charalampos Papamanthou, and Roberto Tamassia. Effective Visualization of File System Access-Control. VizSec 2008 (LNCS 5210).

Monitoramento Visual de Ataques



Fabian Fischer, Florian Mansmann, Daniel A. Keim, Stephan Pietzko, and Marcel Waldvogel. Large-Scale Network Monitoring for Visual Analysis of Attacks. VizSec 2008 (LNCS 5210)

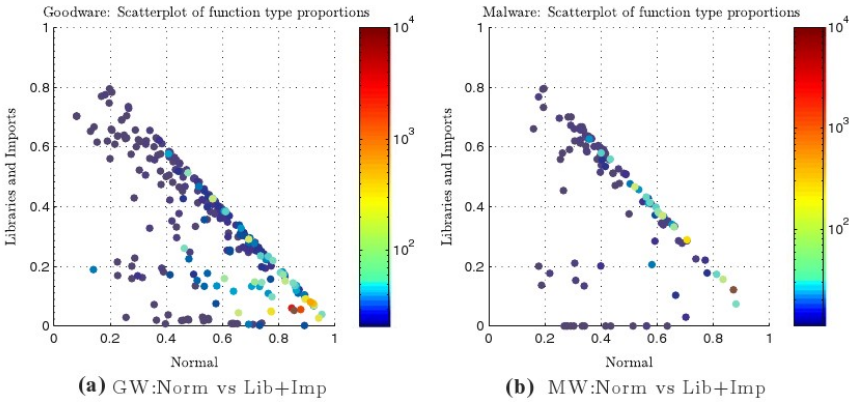
Filogenia de Malware



VX Heavens	Norton	McAfee	ClamAV
IWorm.Mydoom.q	W32.Mydoom.gen@mm	Mydoom.gen@mm	Worm.Mydoom.Sunp
IWorm.Mydoom.u	**not detected**	Mydoom.gen@mm	Worm.Mydoom.Genunp
IWorm.Mydoom.g	W32.Mydoom.G@mm	Mydoom.gen@mm	Worm.Mydoom.Genunp
Win32.Elkern.a	**W32.ElKern.gen**	**W32.ElKern.cav.a**	Worm.Klez.E
IWorm.Klez.a	W32.Klez.A@mm	Klez.worm.gen	Worm.Klez.E
IWorm.Klez.i	W32.Klez.H@mm	Klez.i@mm	Worm.Klez.H
IWorm.Klez.f	W32.Klez.gen@mm	Klez.f@mm	Worm.Klez.E
IWorm.Klez.e	W32.Klez.E@mm	Klez.e@mm	Worm.Klez.E
SpecimenC *	**W32.Elkern.4926**	**W32.Elkern.cav.c**	Worm.Bagle.Gendll
IWorm.Bagle.a I	W32.Beagle.AO@mm	Bagle.dll.dr	Worm.Bagle.AI
SpecimenA *	**not detected**	Bagle.gen@mm	**Trojan.Spamtool.Small.F**
SpecimenB *	W32.Beagle.gen	**not detected**	**Trojan.Spamtool.Small.F**
IWorm.Bagle.s	W32.Beagle.U@mm	Bagle.gen.b@mm	Worm.Bagle.Gendll
IWorm.Bagle.a	W32.Beagle.A@mm	Bagle.a@mm	Worm.Bagle.Gendll
IWorm.Bagle.i	W32.Beagle.J@mm	Bagle.j@mm	Worm.Bagle.Gendll
IWorm.Bagle.j	W32.Beagle.gen	Bagle.k@mm	Worm.Bagle.Kunp
Win32.Alcaul.e	W95.Flee	***Feeling.824intd***	W32.Alcaul.E
Win32.Alcaul.c	**W32.Deimos.1255**	**Deimos**	W32.NgVck.D1
Win32.Alcaul.f	**W95.Nerhook.intd**	**Feeling.1195intd**	W32.NgVck.D2

Md. Enamul Karim, Andrew Walenstein, Arun Lakhotia, and Laxmi Parida. Malware phylogeny generation using permutations of code. Journal of Computer Virology, 1(1):13–23, 2005.

Callgraphs



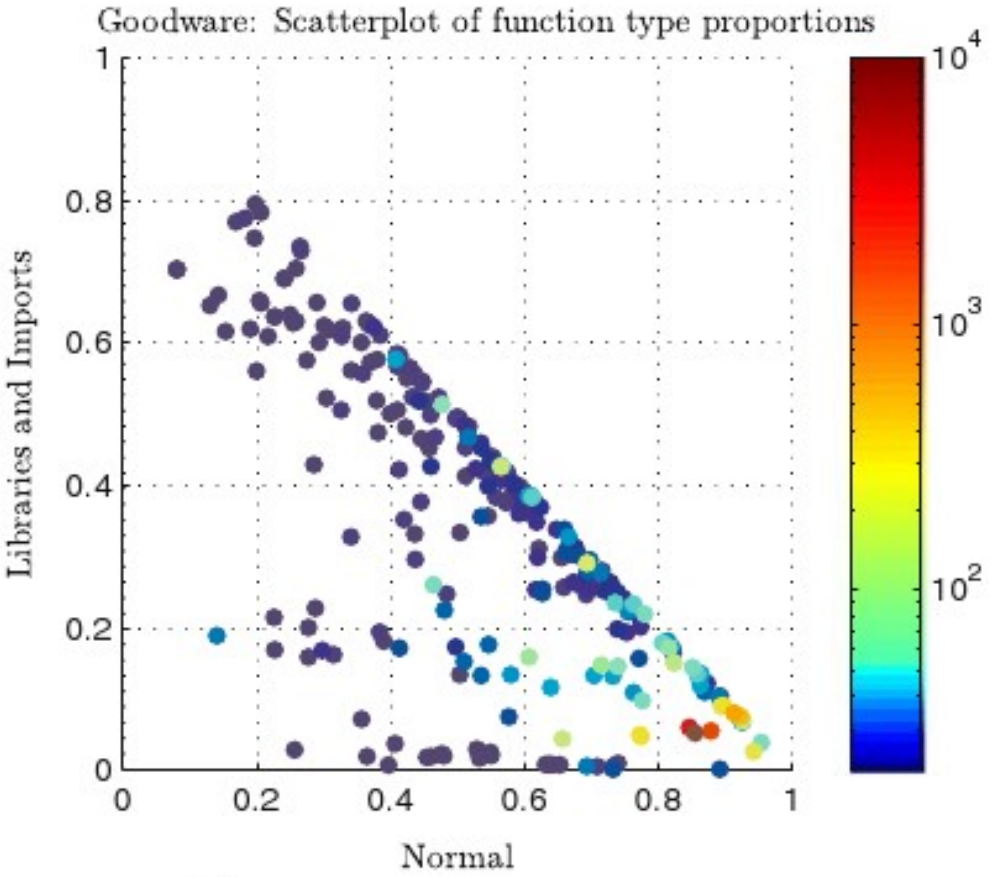
(a) GW:Norm vs Lib+Imp (b) MW:Norm vs Lib+Imp

(c) GW:Norm vs Thunk (d) MW:Norm vs Thunk

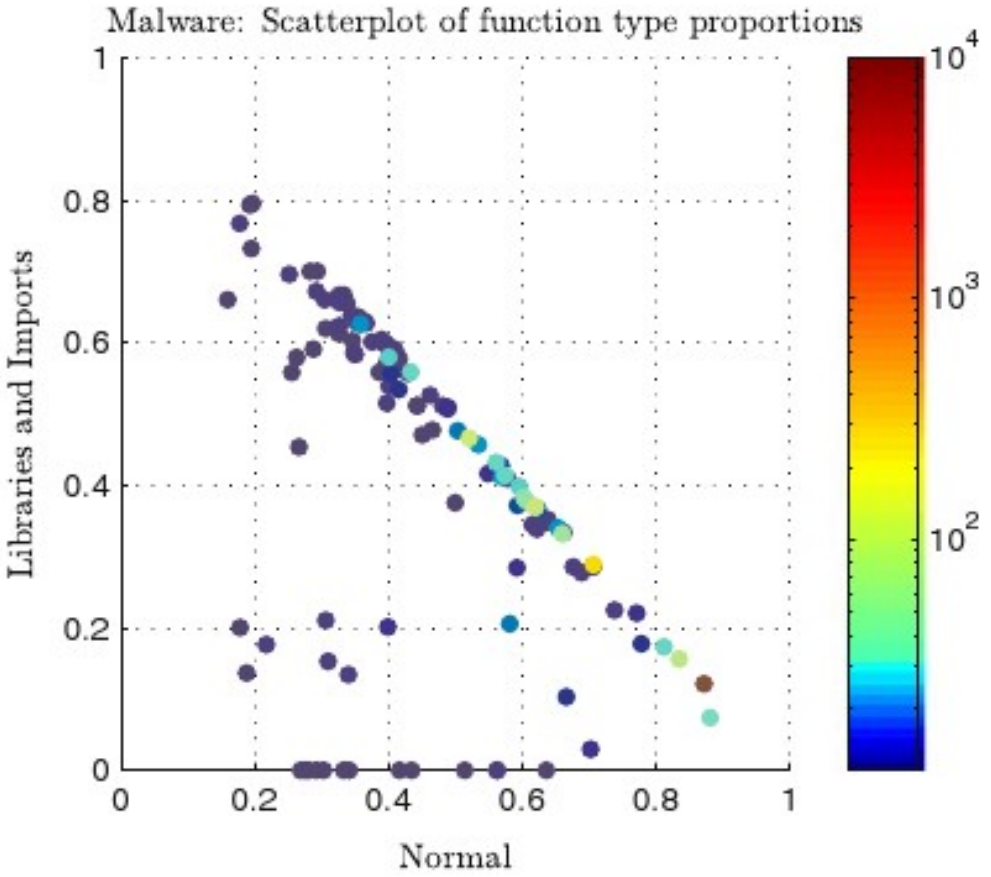
(e) GW:Thunk vs Lib+Imp (f) MW:Thunk vs Lib+Imp

Daniel Bilar. On callgraphs and generative mechanisms. Journal of Computer Virology, 3(4):163–186, 2007.

Callgraphs



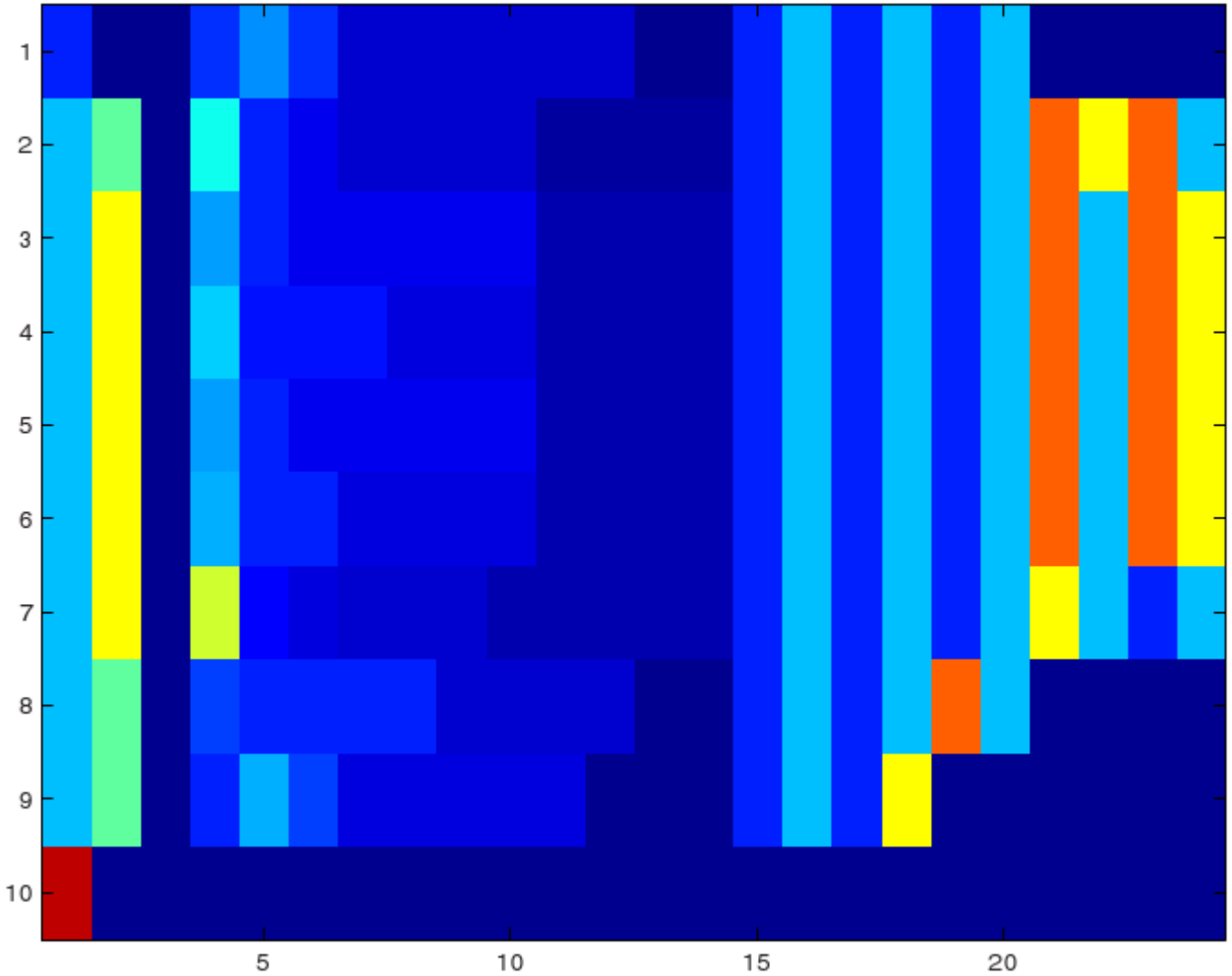
(a) GW:Norm vs Lib+Imp



(b) MW:Norm vs Lib+Imp

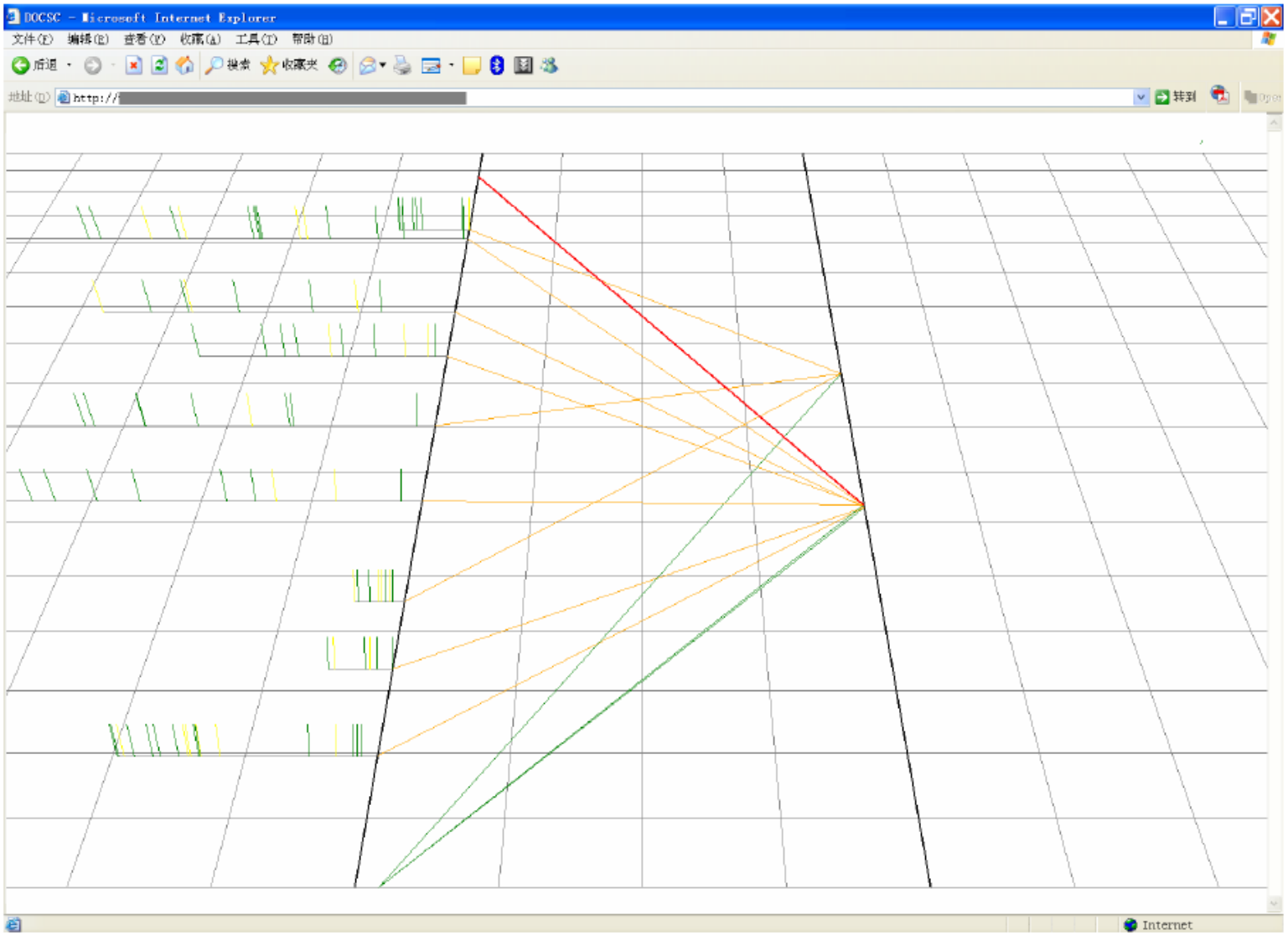
Daniel Bilar. On callgraphs and generative mechanisms. Journal of Computer Virology, 3(4):163–186, 2007.

Visualização de Comportamento de Aplicações



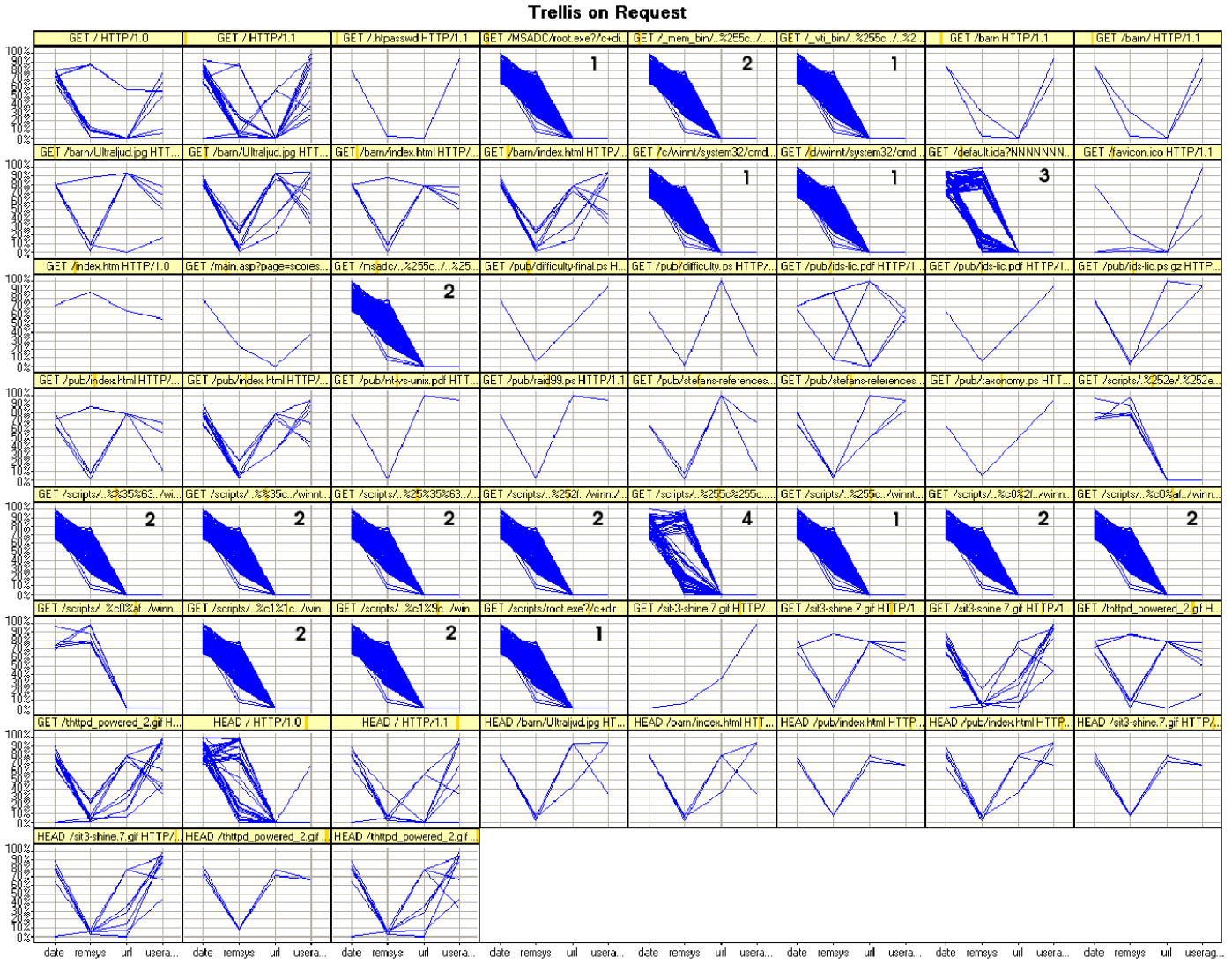
Alessandro Micarelli and Giuseppe Sansonetti. A Case-Based Approach to Anomaly Intrusion Detection. MLDM 2007 (LNCS 4571).

Tráfego Interno de E-mail



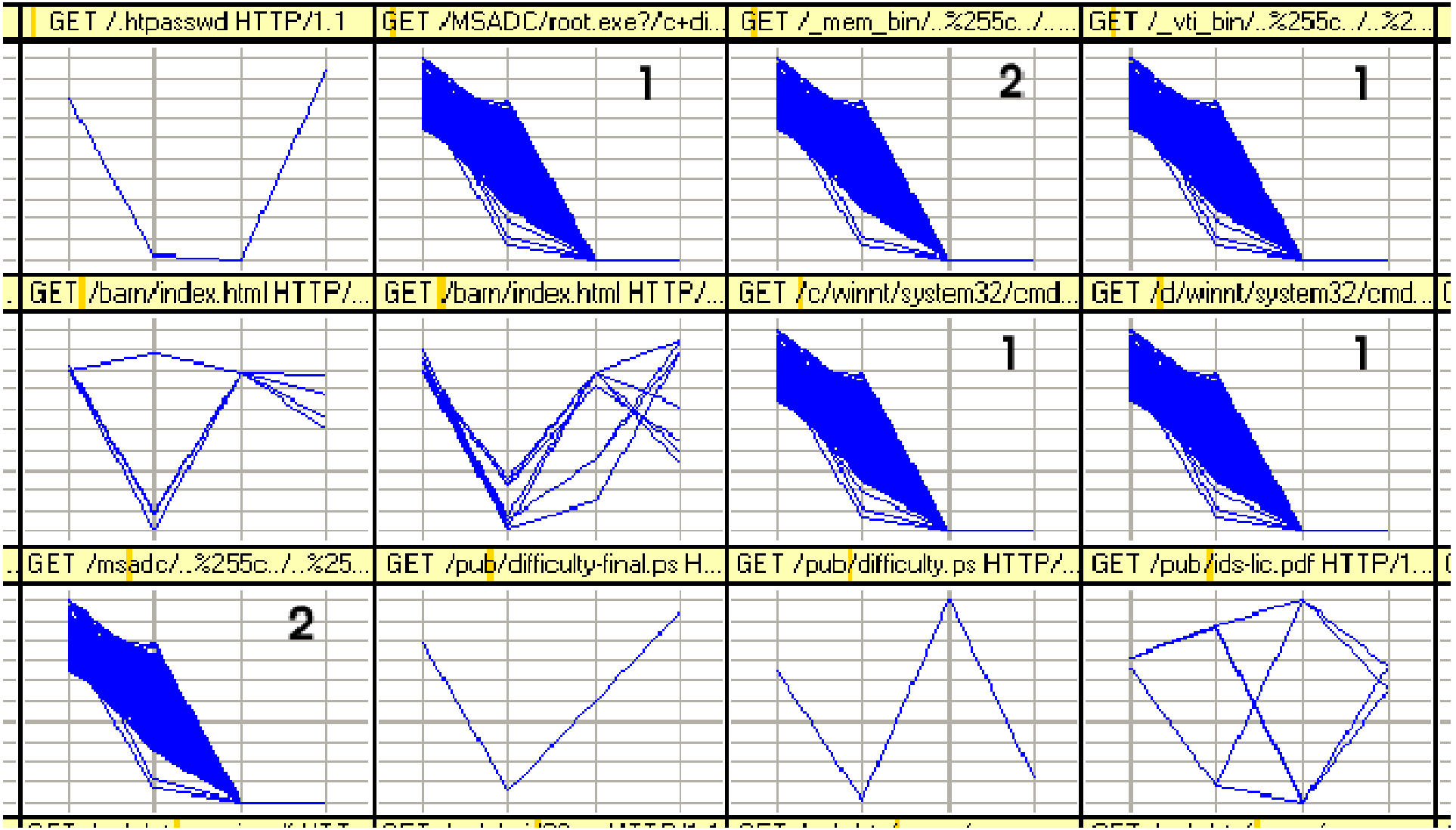
Xiang-Hui Wang and Guo-Yin Zhang. Web-Based Three-Dimension E-Mail Traffic Visualization. APWeb 2006 (LNCS 3842).

Visualização de Acesso a Servidor HTTP



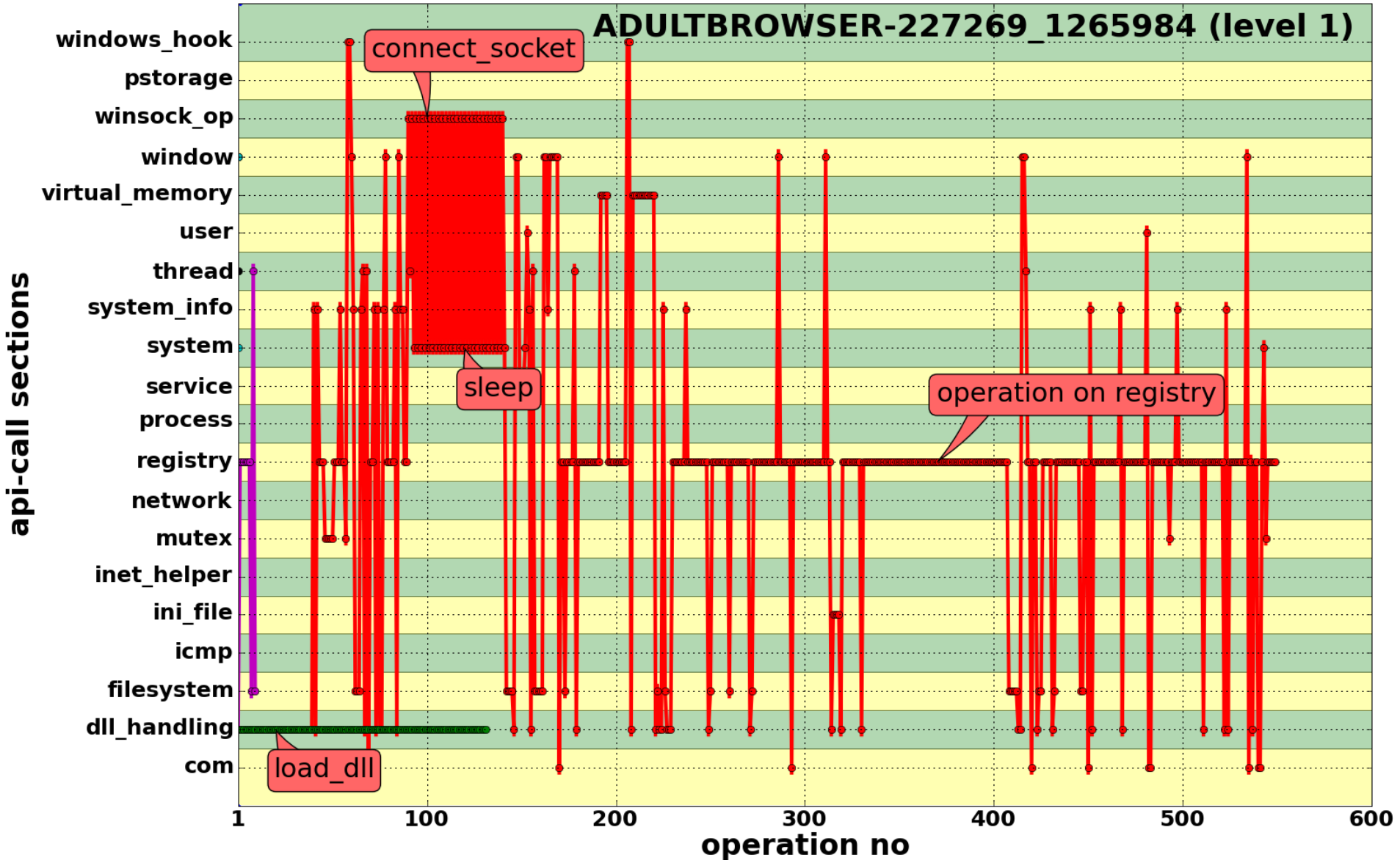
Stefan Axelsson. Visualisation for Intrusion Detection – Hooking the Worm. ESORICS 2003. (LNCS 2808).

Visualização de Acesso a Servidor HTTP



Stefan Axelsson. Visualisation for Intrusion Detection – Hooking the Worm. ESORICS 2003. (LNCS 2808).

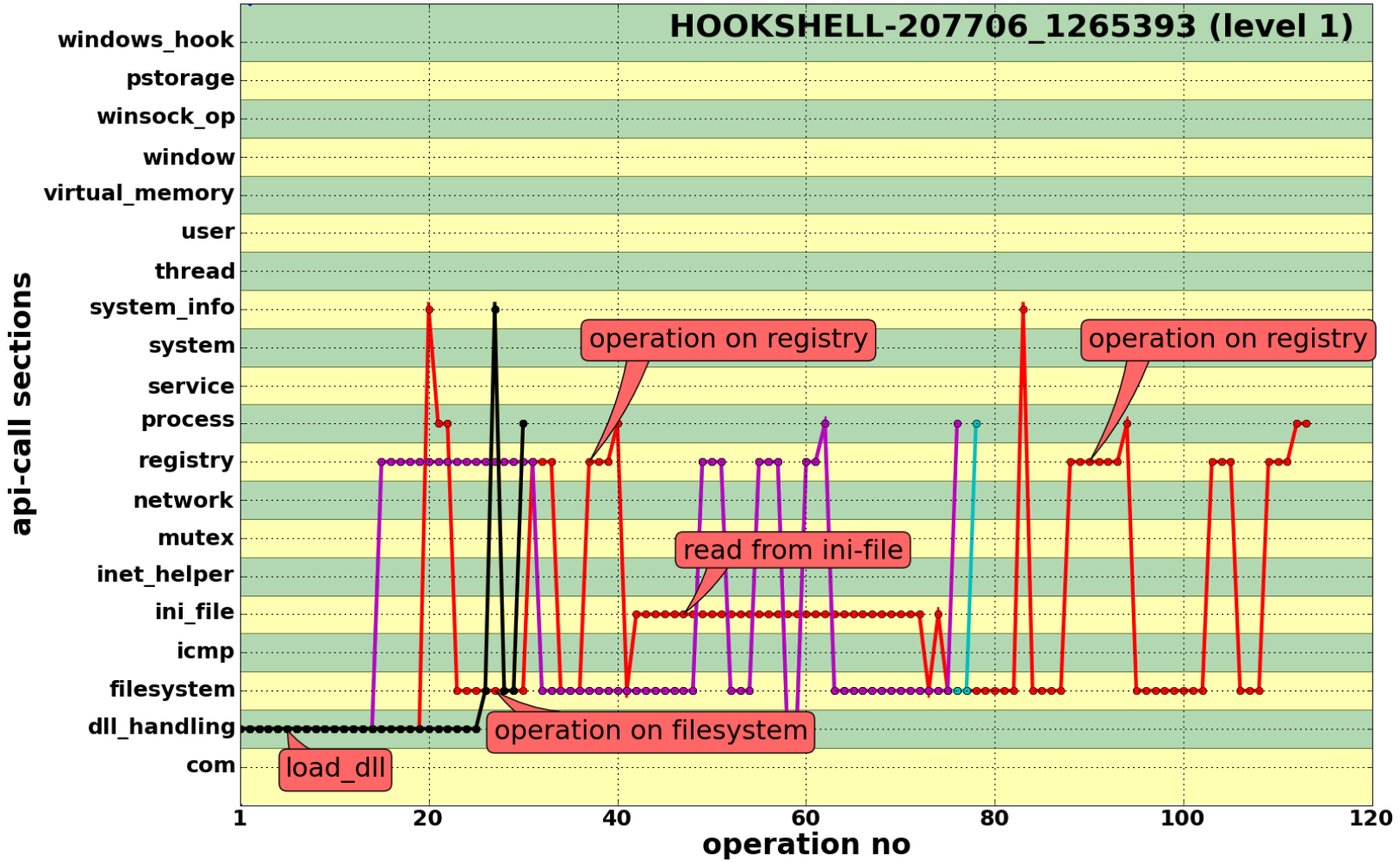
Thread Graphs for Visualizing Malware Behavior - 1



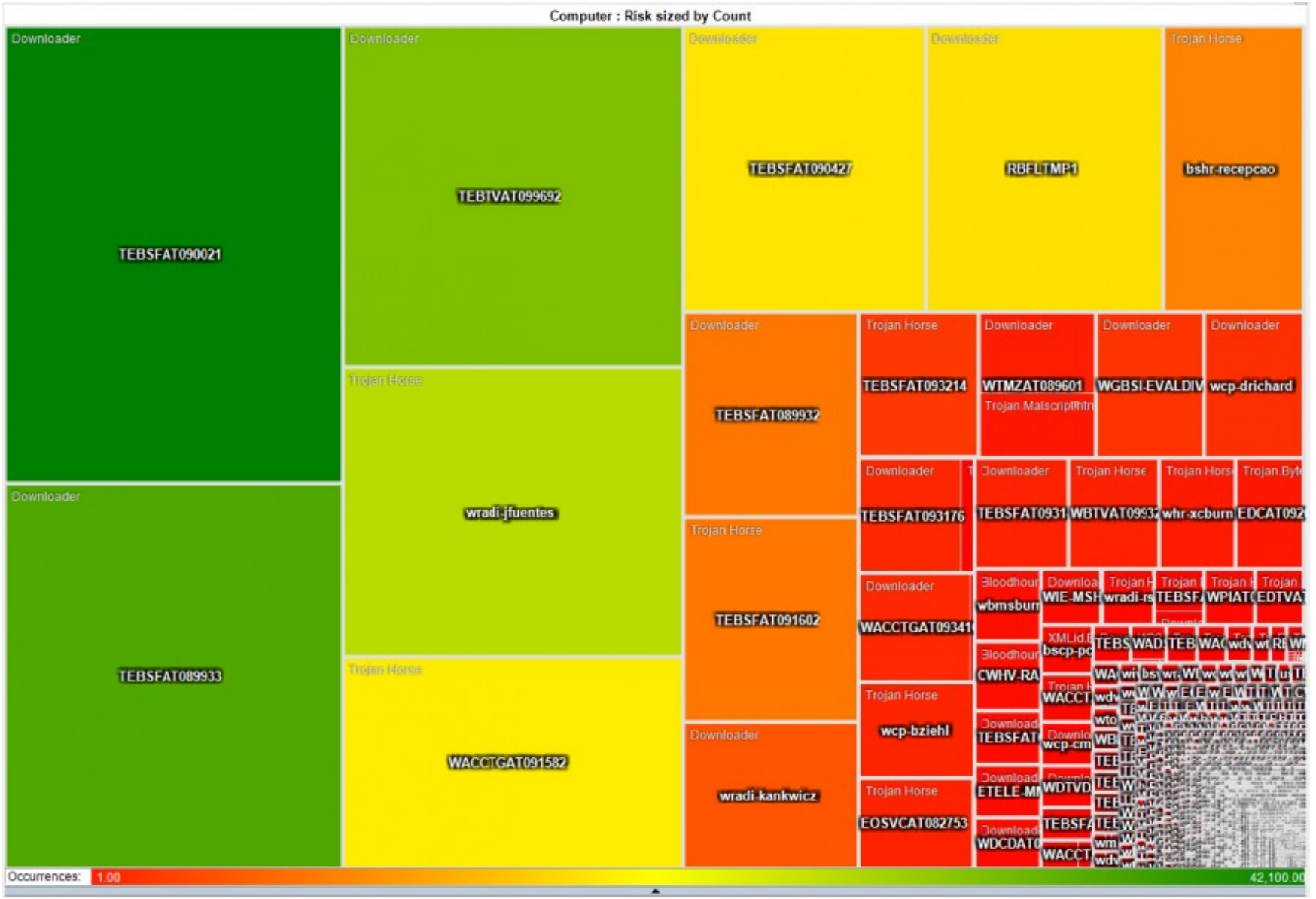
http://honeyblog.org/archives/34-Thread-Graphs-for-Visualizing-Malware-Behavior.html

Thread Graphs for Visualizing Malware Behavior - 2

- Eixo X: é a sequência de ações realizadas durante o tempo;
- Eixo Y: indica a operação/seção da ação efetuada.



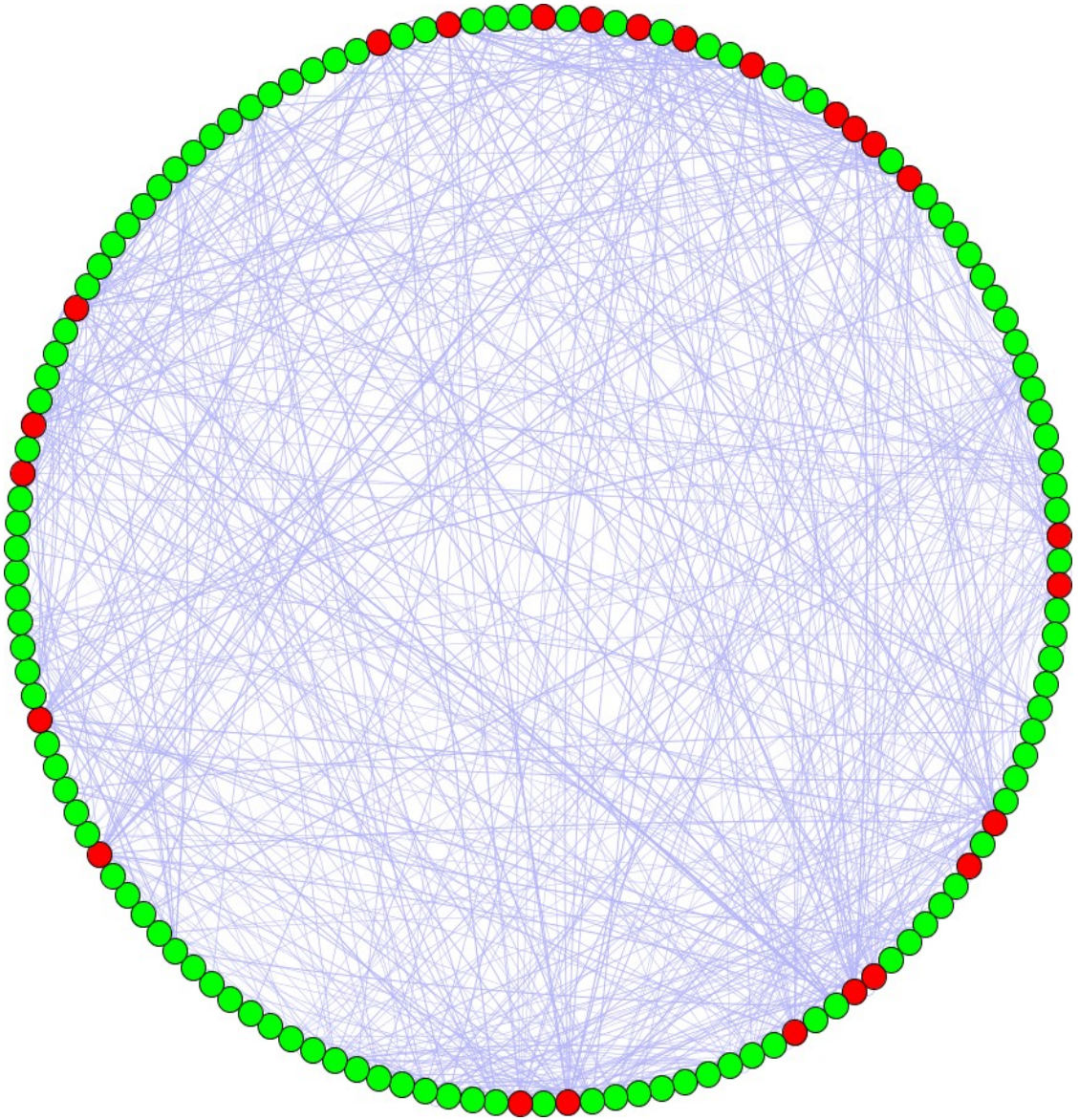
AV Malware Detect Heat Map



http://secviz.org/content/av-malware-detect-heat-map

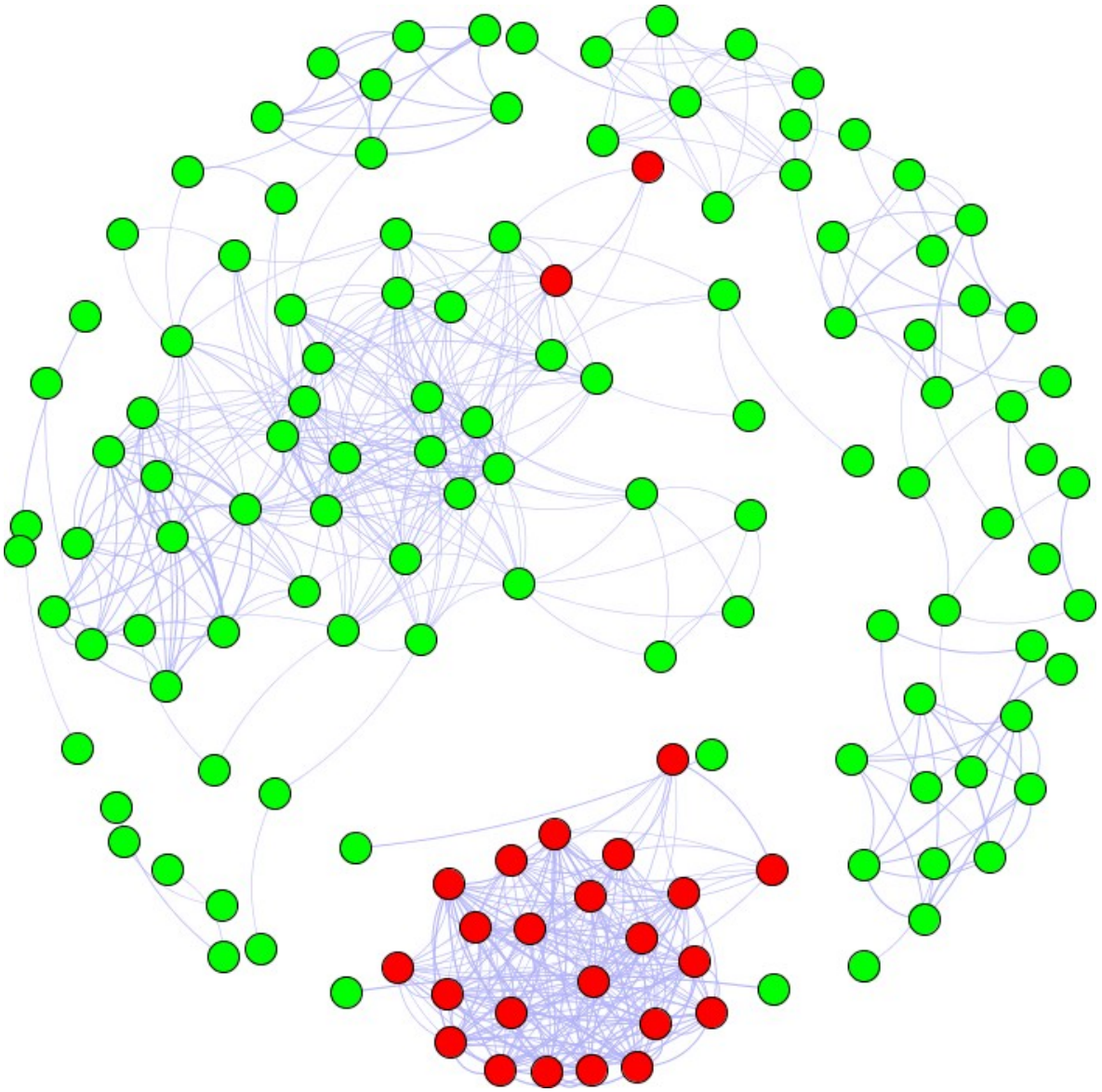
Software

Software



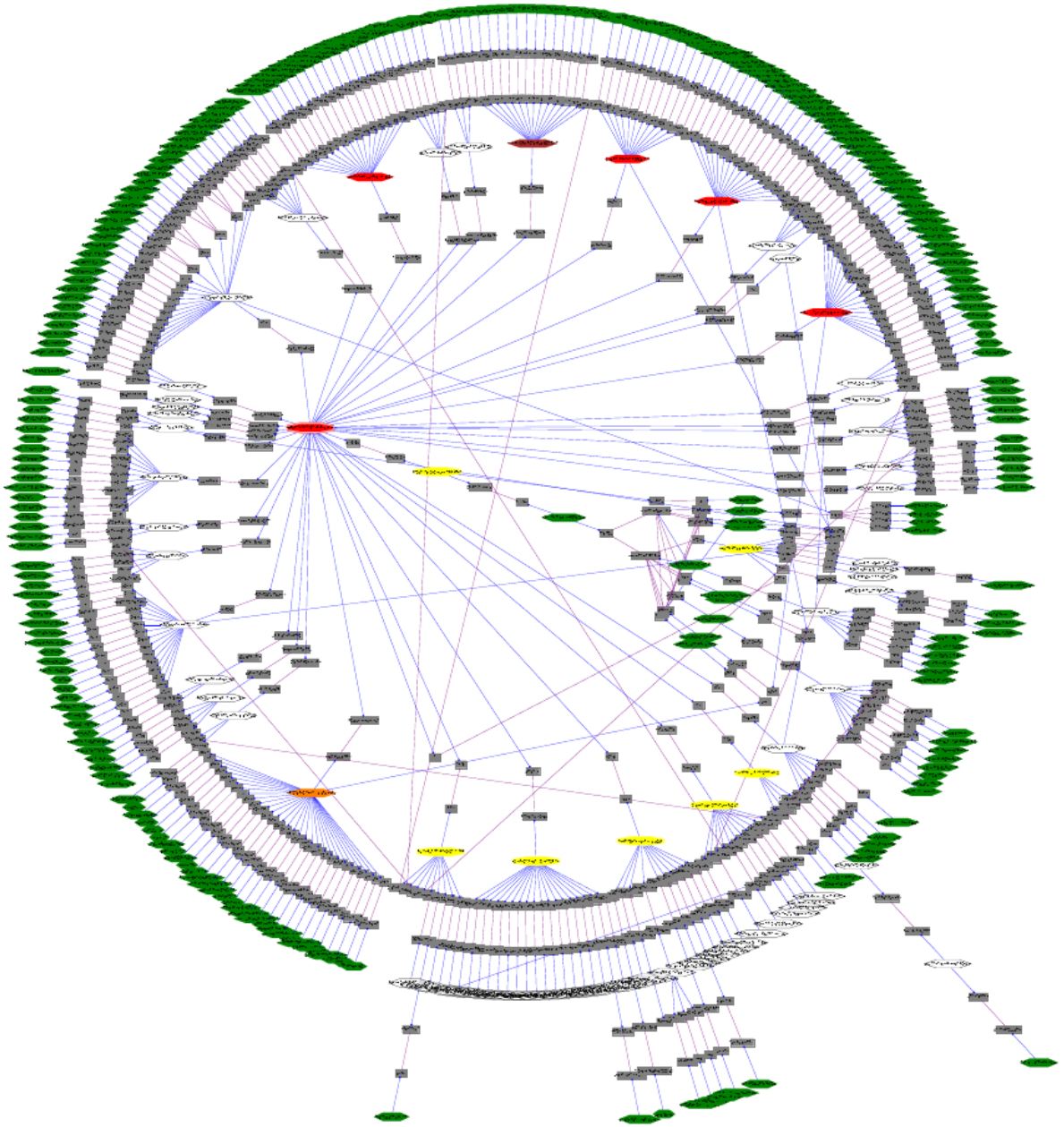
JUNG (*Java Universal Network/Graph Framework*): visualização de similaridade entre malware.

Software



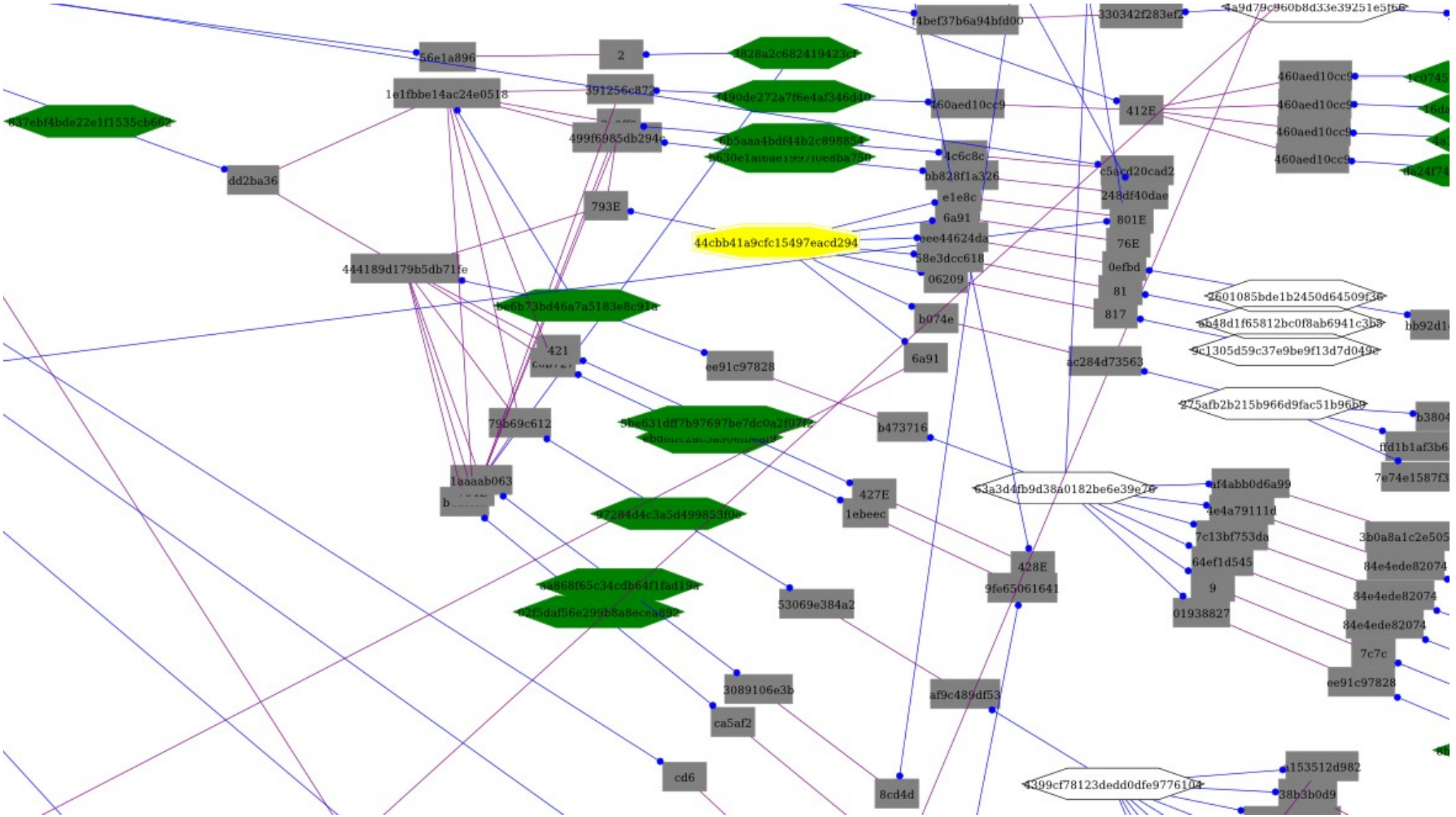
JUNG (*Java Universal Network/Graph Framework*): visualização de similaridade entre malware.

Software



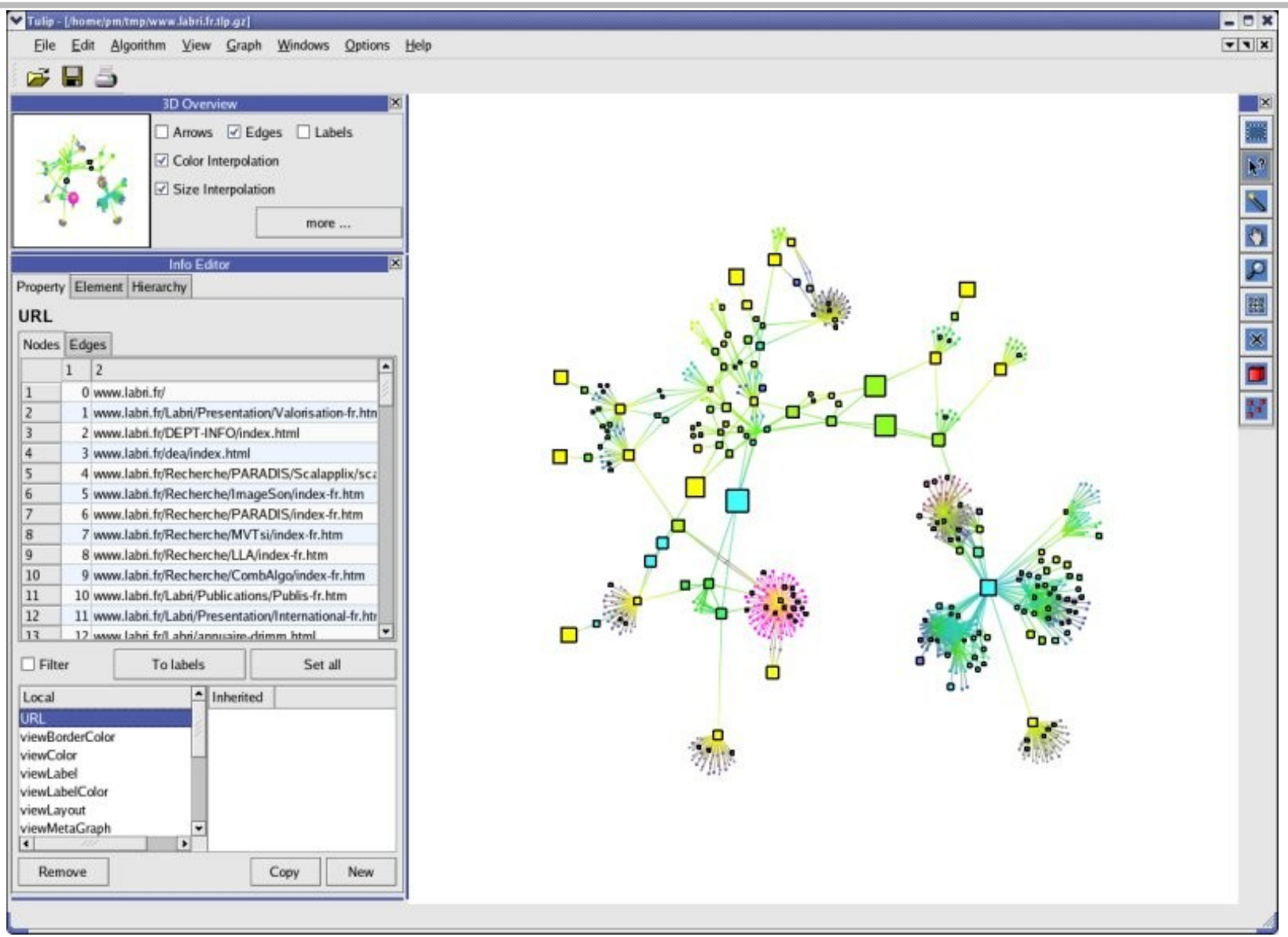
Graphiviz: visualização de 300 sites em 40 países.

Software



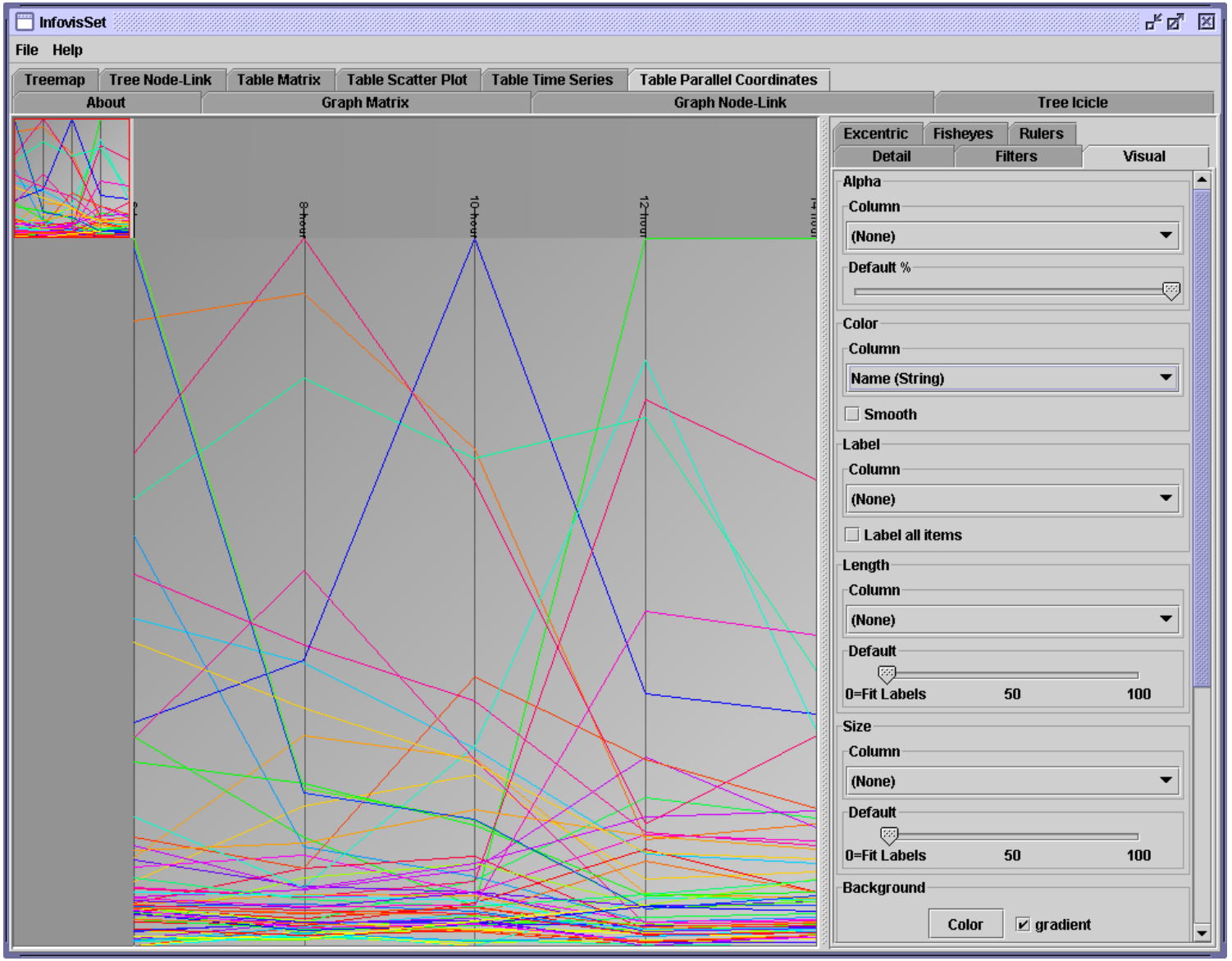
Graphiviz: visualização de 300 sites em 40 países. http://www.graphviz.org/Gallery/twopi/twopi2.html

Software



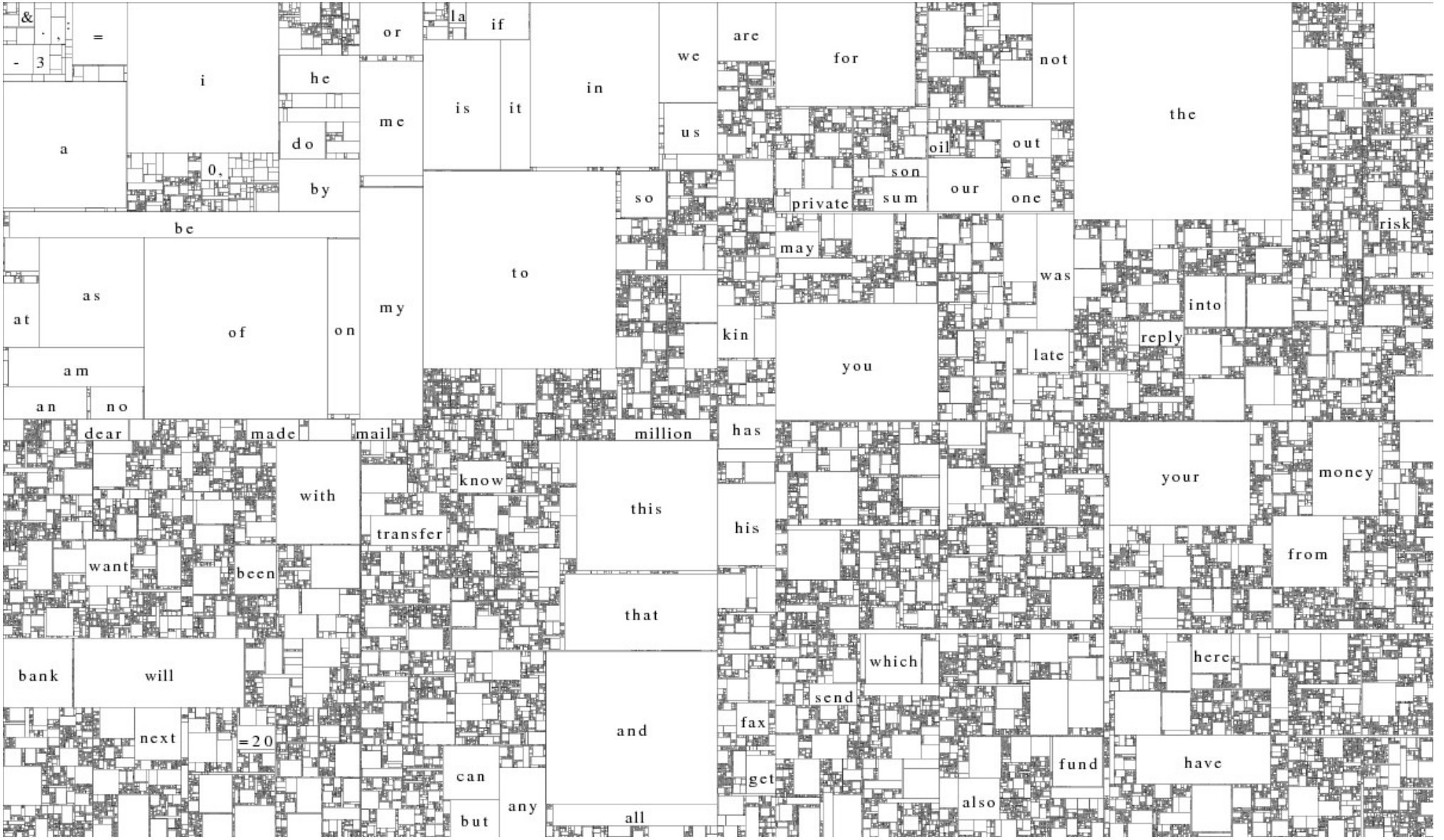
Tulip (tulip-software.org): Estrutura dos arquivos em um servidor HTTP.

Software



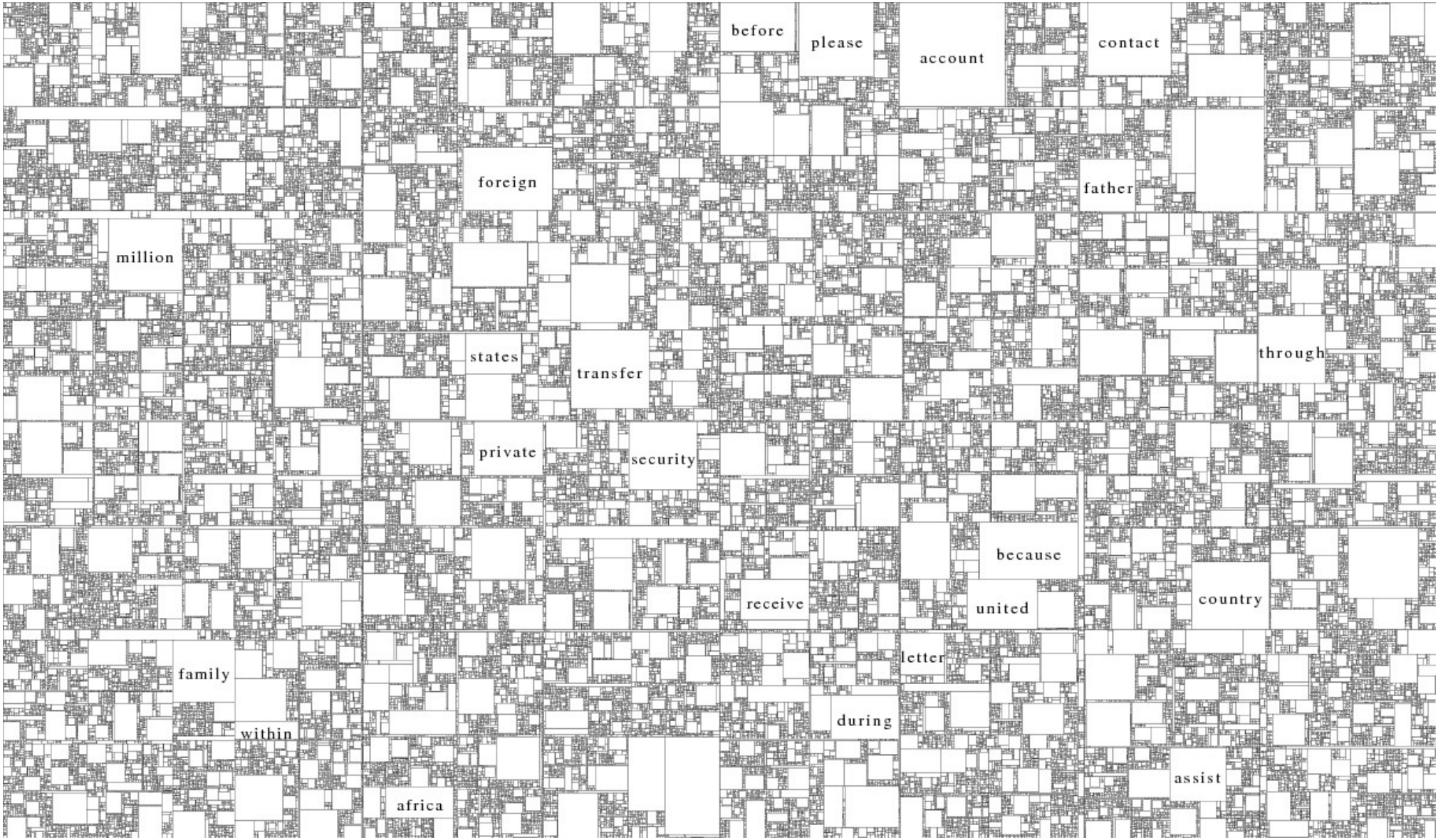
InfoVis Toolkit (ivtk.sourceforge.net): coordenadas paralelas anotadas.

Software



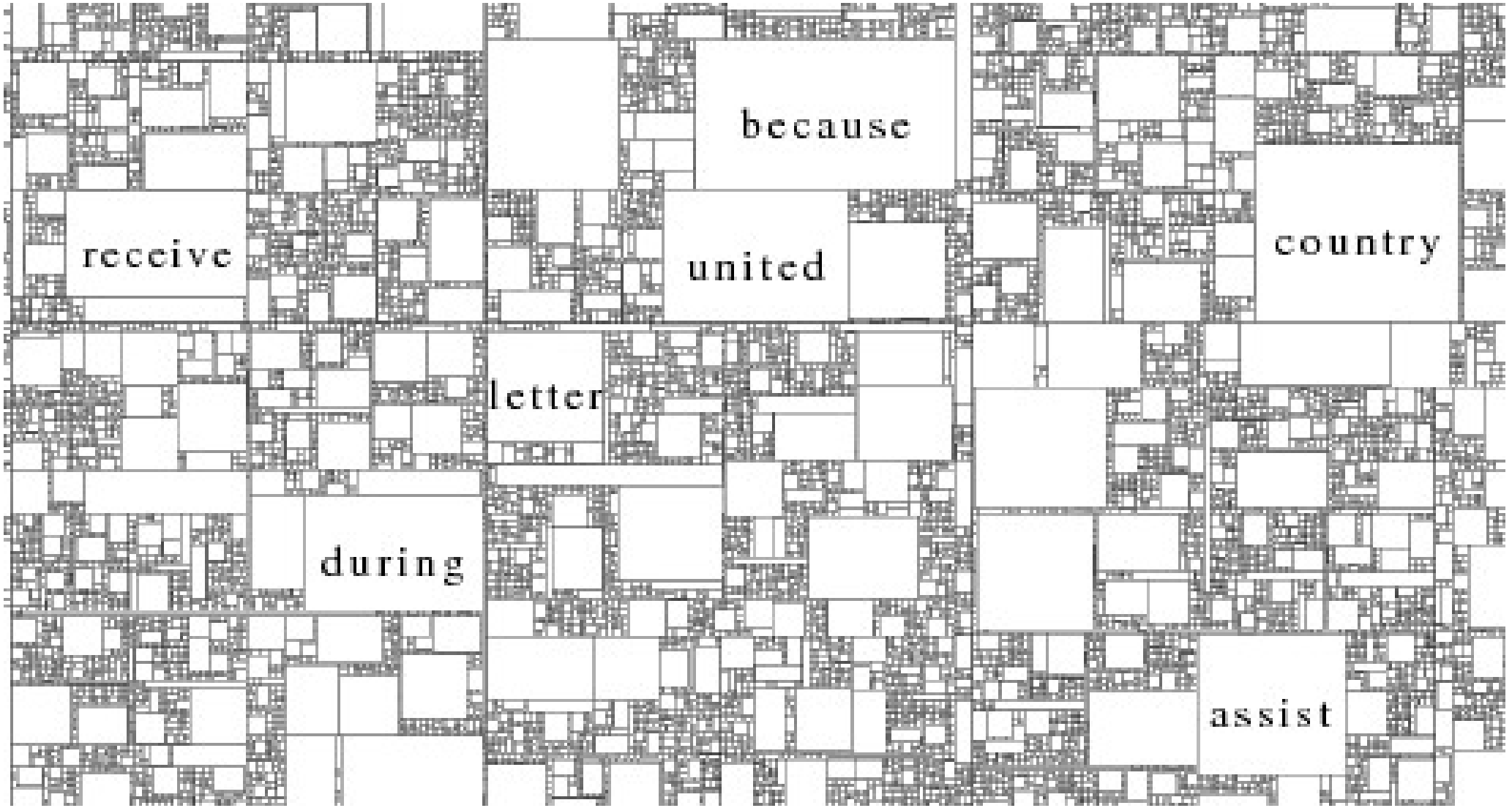
Processing (processing.org): Treemap de frequências de palavras em *scams* tipo 419.

Software



Processing (processing.org): Treemap de frequências de palavras em *scams* tipo 419.

Software



Processing (processing.org): Treemap de frequências de palavras em *scams* tipo 419.

Sugestões para Estudos Complementares

Estudos Complementares

- Eventos:
  - SBSEG, Sibgrapi.
  - VizSec / IEEE VisWeek (www.vizsec.org)
  - IEEE International Conference
- Exemplos:
  - DAVIX Live CD (davix.secviz.org)
  - Honeyblog.org
  - www.honeynet.org.au