



UNIVERSIDADE DE AVEIRO

DEPARTAMENTO DE ELECTRÓNICA, TELECOMUNICAÇÕES E
INFORMÁTICA

DESEMPENHO E DIMENSIONAMENTO DE REDES

Network Statistical Analysis

8240 - MESTRADO INTEGRADO EM ENGENHARIA DE
COMPUTADORES E TELEMÁTICA

Bernardo Ferreira
NMec: 67413 | P4G1

Bruno Silva
NMec: 68535 | P4G1

Docente: Susana Sargento

Abril de 2016
2015-2016

Conteúdos

1	Introdução	2
2	Exercício 1	3
3	Exercício 2	4
4	Exercício 3	5
5	Exercício 4	6
6	Exercício 5	8
7	Exercício 6	10
8	Exercício 7	11
9	Exercício 8	12
10	Exercício 9	14
11	Exercício 10	17
12	Exercício 11	18
13	Exercício 12	20
14	Exercício 13	22
15	Exercício 14	23
16	Exercício 15	25
17	Exercício 16	27
18	Exercício 17	31
19	Exercício 18	31
20	Exercício 19	37
21	Exercício 20	42
22	Conclusão	45

1 Introdução

Este relatório vem explicar as conclusões retiradas da observação e compreensão das diferentes métricas e estatísticas pedidas para calcular e apresentar neste segundo guião prático.

Este relatório vai ser estruturado da mesma forma que o guião, seguindo a mesma numeração dos exercícios.

No ficheiro de código "*baseStats1.py*" os exercícios foram executados um de cada vez, comentando todos exceto o que queríamos executar. Desta forma evitamos que para executar o ultimo exercício tenhamos de ver todos os outros, mais ainda haviam exercícios que necessitavam de mais que um gráfico pelo que foram atribuídos, na maior parte dos casos, números sequenciais o que iria sobrepor com o gráfico do exercício seguinte desta forma impossibilitando descomentar o código todo e executar.

2 Exercício 1

No exercício 1 era pedido a análise dos perfis de download para 40 serviços.

Após serem carregados os perfis de tráfego dos 40 serviços é possível observar que todos os serviços se enquadram em um de dois grupos. Um primeiro grupo onde o tráfego é mais periódico e com maior volume, e um segundo serviço onde o tráfego é mais irregular e com menor volume. A visualização gráfica destes dois grupos está representada nas figuras em baixo, usando por simplificação apenas um serviço de cada grupo, sendo que a figura 1 representa os serviços do primeiro grupo e a figura 2 representa os serviços do grupo 2.

Olhando para os serviços e tentando obter um conhecimento sobre o tipo de serviço associado pode-se inferir que o primeiro grupo representa um volume e uma distribuição de pacotes semelhante a um serviço multimédia com streaming de dados. Quando olhamos para o segundo grupo de serviços o tráfego é muito mais característico com o tradicional browsing por páginas de Internet.

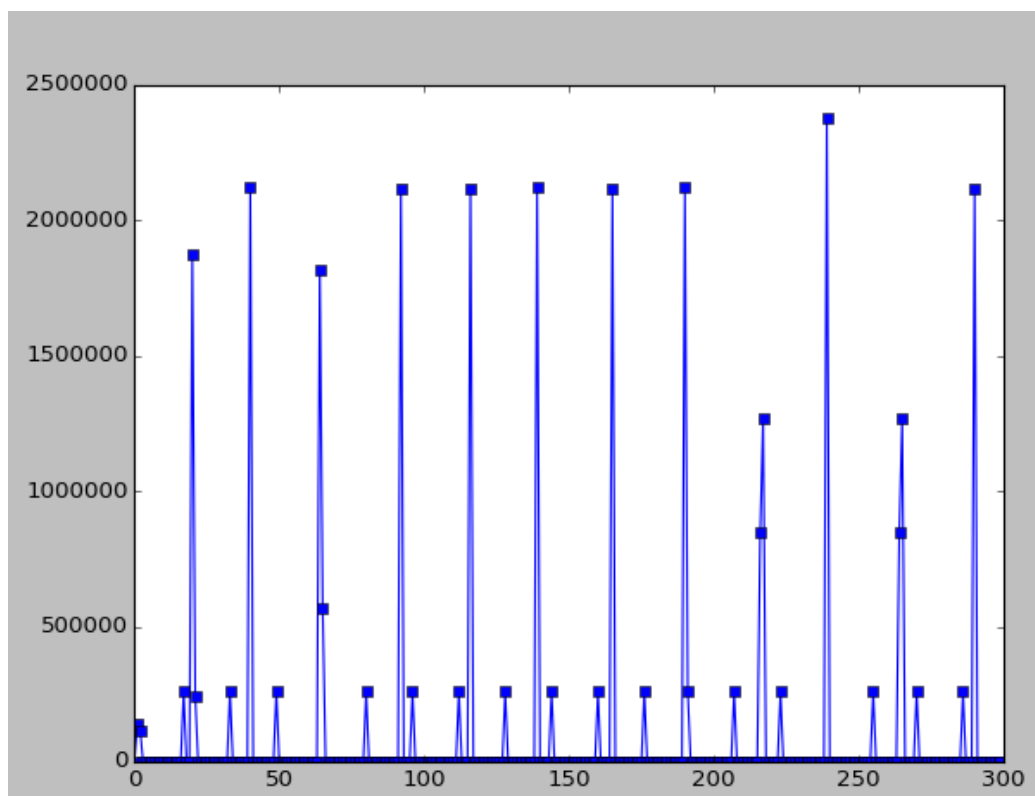


Figura 1: Exemplo de um serviço do primeiro grupo (serviço 0)

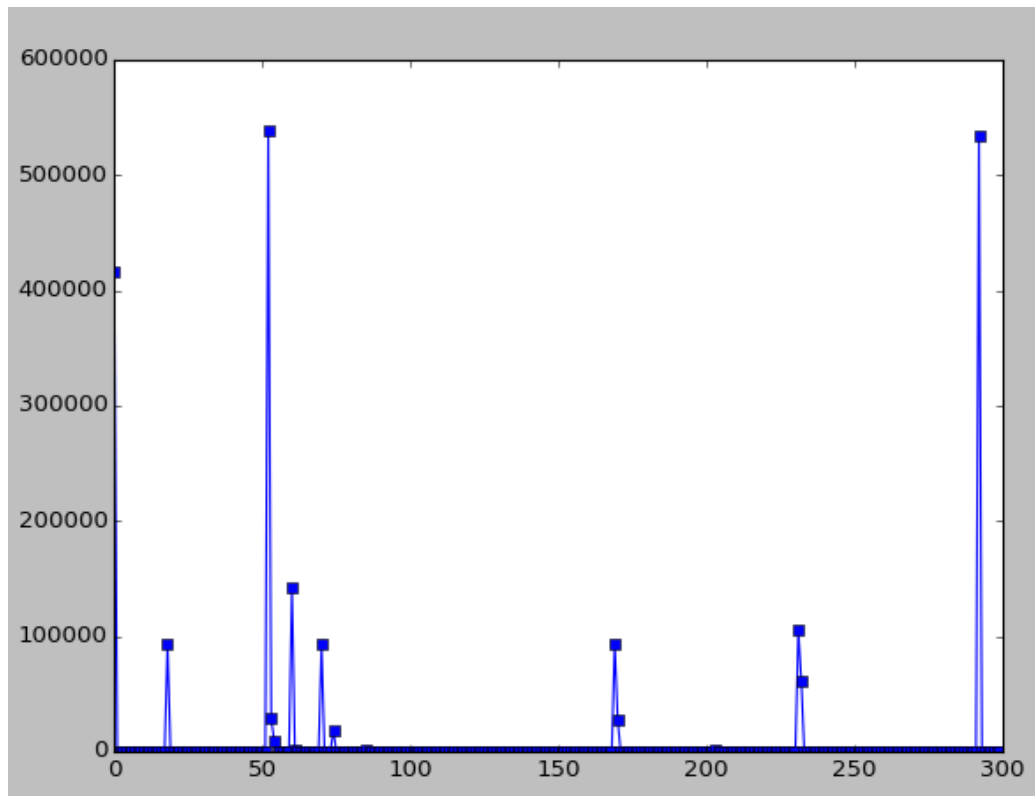


Figura 2: Exemplo de um serviço do segundo grupo (serviço 20)

3 Exercício 2

No exercício 2 era pedido para inferir estatísticas simples sobre os serviços e classificar os serviços em diferentes classes.

A análise dos valores estatísticos vem comprovar o que já tinha sido referido no exercício anterior sobre a existência de dois grupos distintos de serviços em que o primeiro tem um maior valor de pacotes, o que pode ser comprovado com a media ser mais elevada, que o segundo. Continuando a analisar os valores produzidos pode-se concluir que ambos os grupos de serviços durante grande parte do seu tempo, pelo menos metade, não recebem qualquer pacote uma vez que a mediana é 0 para os 40 serviços.

Uma vez que os picos de tráfego na primeira classe de serviços são muito mais elevados do que na segunda como é possível verificar comparando a figura 1 e 2 temos que a variância também vai ser mais elevada nesta primeira classe.

Quando olhamos para os valores de skewness vemos que também são positivos no segundo grupo, isto também é esperado uma vez que a skewness mede a simetria da função de distribuição de probabilidade. Como o segundo grupo de perfis tem muitos pontos a zero mas a média não é zero temos que a distribuição vai ser assimétrica logo a skewness vai ser positiva visto que essa assimetria se encontra para o lado esquerdo do eixo do x.

A kurtosis é positiva no segundo serviço pois como os pontos são mais concentrados à volta de zero o pico vai ser mais elevado nesse ponto. Uma vez que a kurtosis mede o pico ou achatamento da função de distribuição de probabilidade faz sentido que tenha valores positivos.

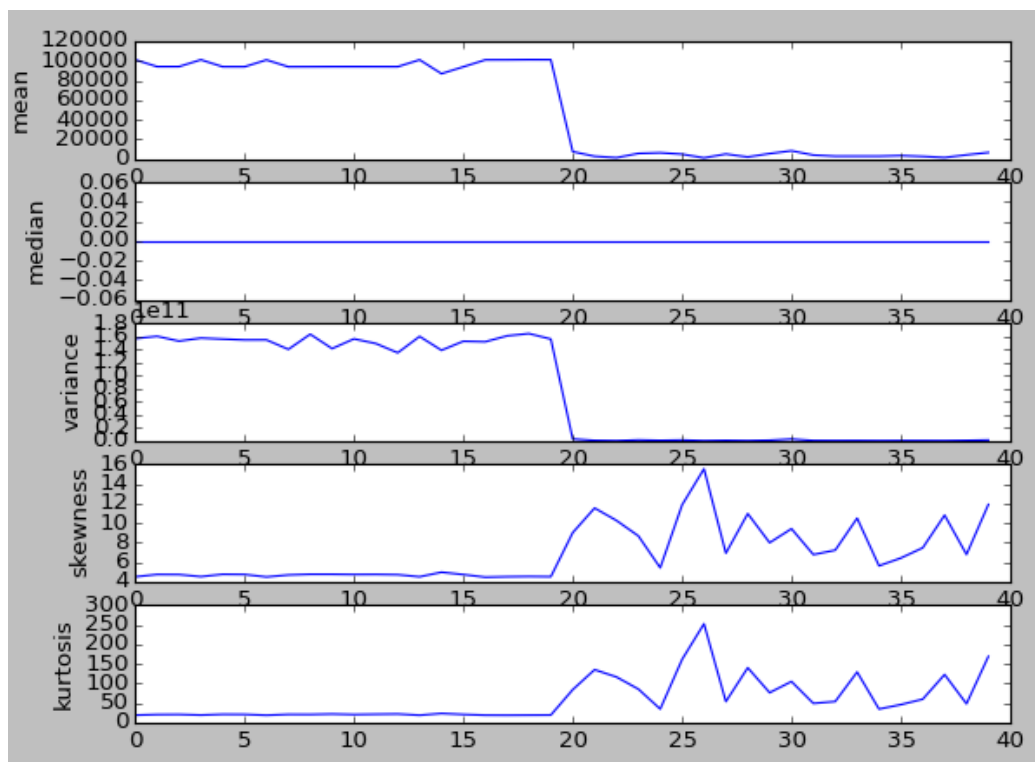


Figura 3: Análise estatística para os 40 serviços agrupada num gráfico

4 Exercício 3

No exercício 3 era pedido para analisar a nossa captura de pacotes do Youtube e comparar com os grupos de serviços observados anteriormente.

Observando a figura 4 é possível identificar um tráfego por burst mas com picos em que são recebidos grandes quantidades de pacotes.

A nível de comparação com a figura 1 e 2 é possível identificar o tráfego do Youtube como pertencendo ao primeiro grupo, o que faz sentido uma vez que concluímos, no exercício 1, que o primeiro grupo se tratava de tráfego multimédia.

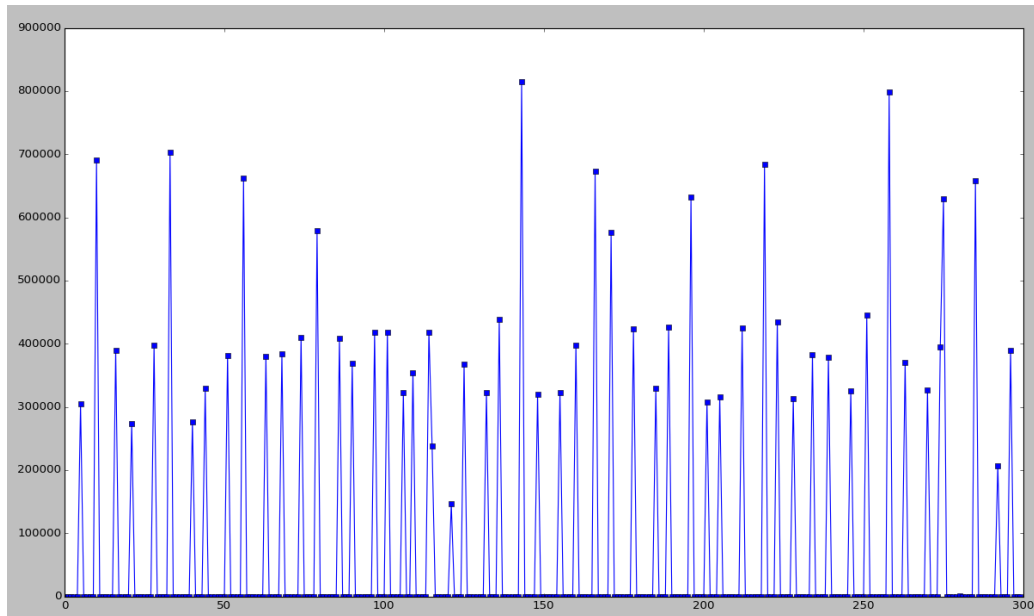


Figura 4: Gráfico do tráfego gerado pela captura de pacotes do Youtube

5 Exercício 4

No exercício 4 era pedido para calcular e analisar a função de densidade de probabilidade(PDF) e a função de densidade cumulativa(CDF) para todos os serviços.

Ao analisar a PDF dos serviços de 0-19(azul), observamos que as probabilidades estão mais distribuídas pelo eixo dos xx, devido aos pacotes serem recebidos em bursts, enquanto que nos serviços de 20-39(vermelho) há grande probabilidade de obter zero pacotes devido a haver longos períodos em que não há receção de qualquer pacote. Este facto também justifica os valores mais elevados de kurtosis e skewness observados nestes serviços, em relação aos serviços de 0-19.

Confirmando as conclusões que já tínhamos inferido anteriormente, observa-se que o nosso perfil do YouTube se aproxima mais dos serviços de 0-19, tal como é possível observar nas figuras 5 e 6.

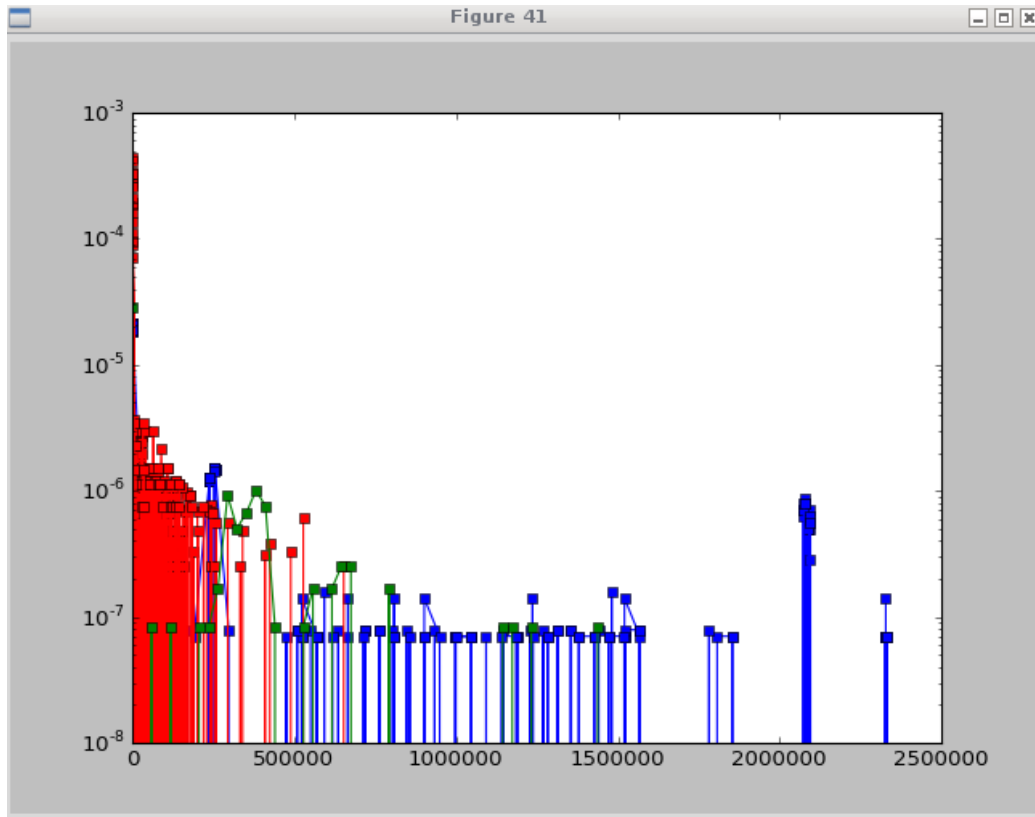


Figura 5: PDF dos perfis de 0-19(azul), 20-39(vermelho) e YouTube(verde)

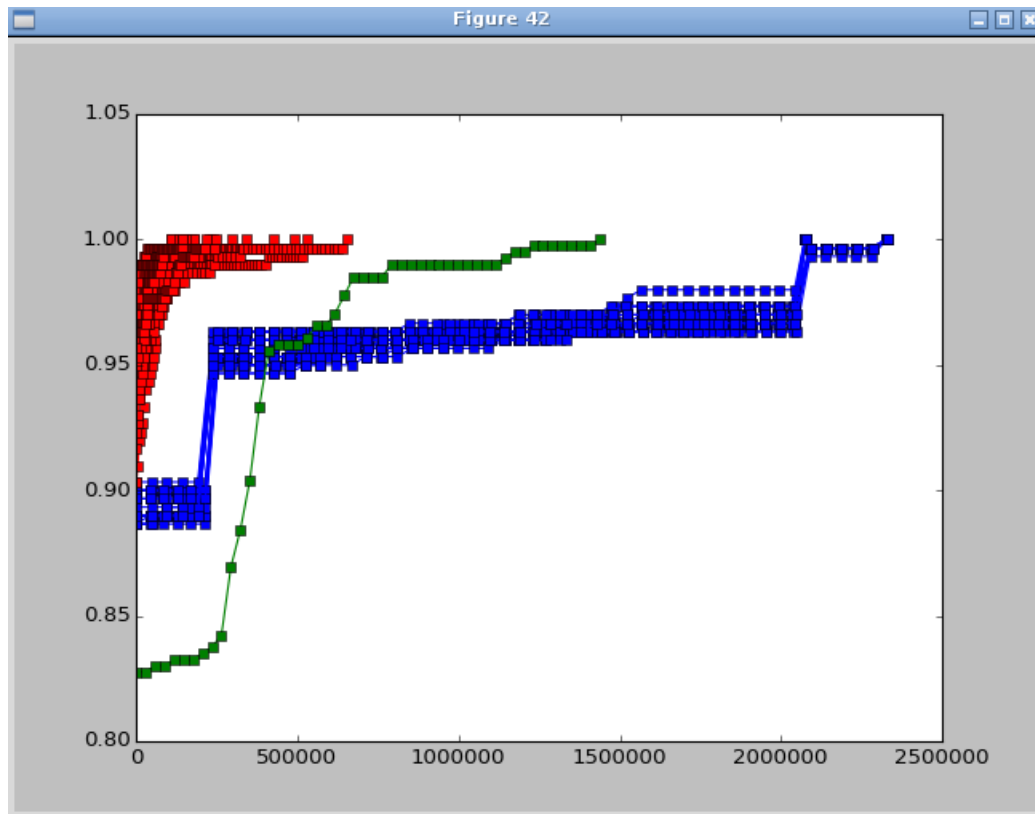


Figura 6: CDF dos perfis de 0-19(azul), 20-39(vermelho) e YouTube(verde)

6 Exercício 5

No exercício 5 era pedido para calcular e analisar os gráficos de P-P e Q-Q entre os diferentes perfis e também entre o nosso perfil do YouTube.

Analisando as figuras 7 e 8 e sabendo que os pontos quanto mais próximos da linha estiverem, maiores são as semelhanças, podemos concluir mais uma vez que há grandes diferenças entre os dois grupos de perfis de tráfego disponibilizados, já que a dispersão de pontos está bastante afastada da linha vermelha.

Nas figuras 9 e 10, temos o Q-Q e P-P calculados entre o nosso YouTube e os serviços do primeiro grupo(figura 10) e do segundo grupo(figura 9). Neste caso, conclui-se mais uma vez a aproximação do nosso YouTube aos perfis de 0-19(figura 10), já que a distância dos pontos à reta é menor do que o observado com os perfis 20-39(figura 9).

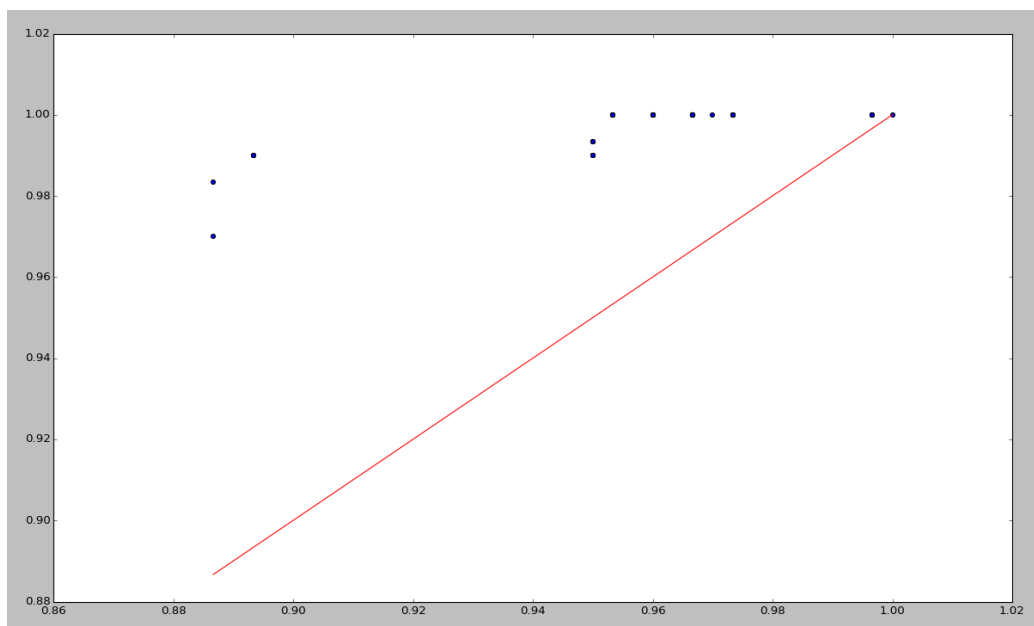


Figura 7: P-P entre os perfis 0(eixo xx) e 20(eixo yy)

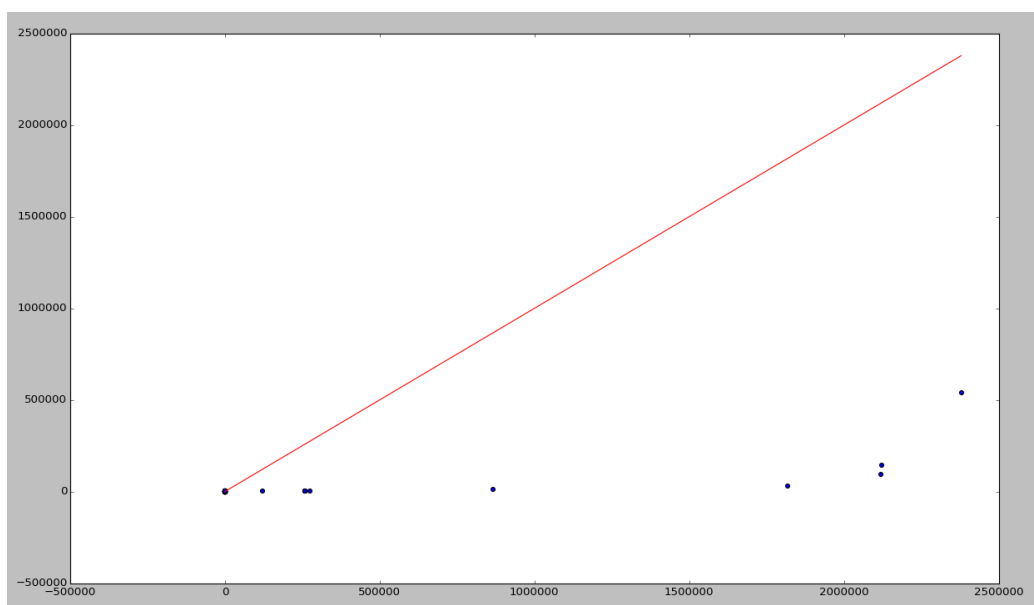


Figura 8: Q-Q entre os perfis 0(eixo xx) e 20(eixo yy)

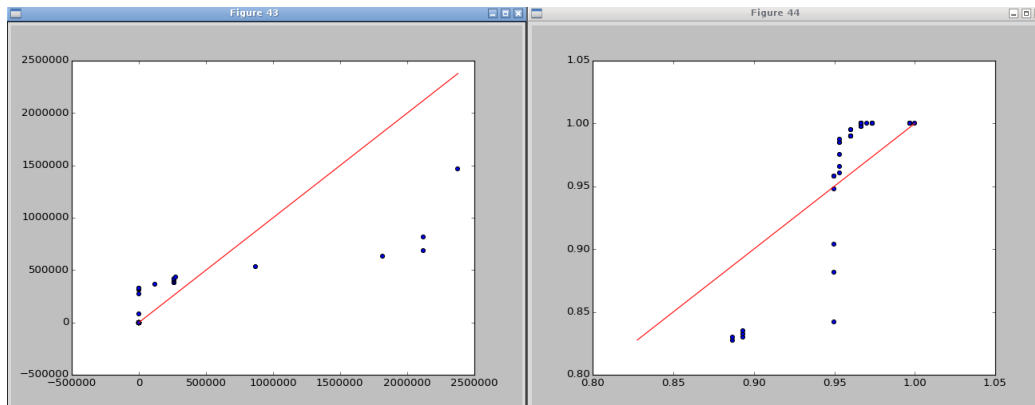


Figura 9: Q-Q e P-P entre o perfil 20 e o nosso YouTube

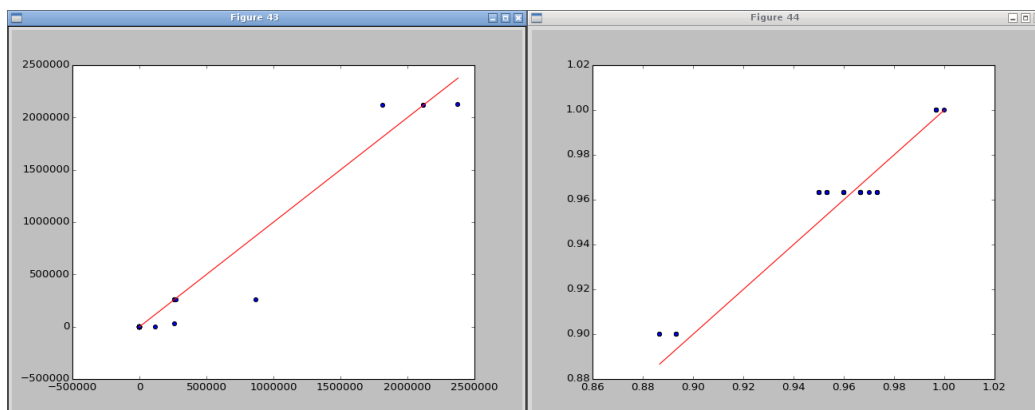


Figura 10: Q-Q e P-P entre o perfil 20 e o nosso YouTube

7 Exercício 6

Neste exercício é pedido para calcular as distancias relativas entre PDF's usando a distancia de Hellinger. Estas distancias foram calculadas e inseridas num gráfico onde é possível de forma muito evidente distinguir os dois diferentes grupos de aplicações que já tinham sido identificados previamente.

Analisando o gráfico a baixo é fácil de concluir que o serviço mais parecido com o Youtube são os serviços de 0-19 uma vez que a distancia entre as pdf's é inferior. Isto é consistente com o que já tínhamos concluído previamente sobre a semelhança do nosso Youtube com os restantes serviços.

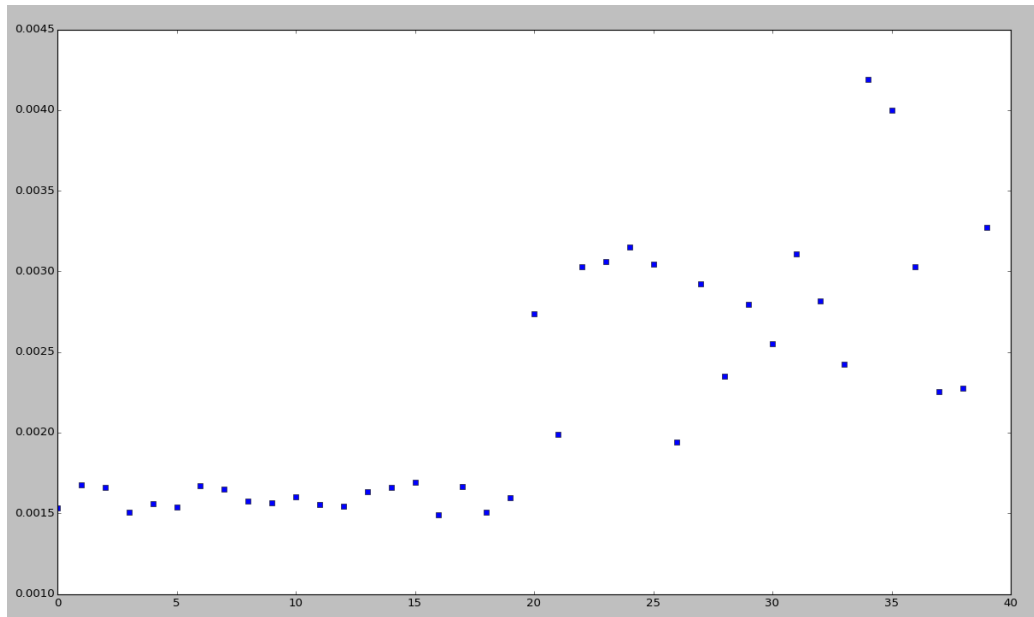


Figura 11: Distancias relativas entre PDF's usando distancias de Hellinger entre o Youtube e os 40 serviços

8 Exercício 7

Neste exercício era pedido para realizar o teste de Kolmogorov–Smirnov entre todos os grupos de perfis.

Como é possível verificar pela figura 12, os valores do D e do p-value entre os perfis 0-1, sustentam a conclusão que se tratam de perfis do mesmo serviço. Em relação à comparação entre o nosso YouTube e o perfil 0 e 30, é possível verificar que não se enquadra totalmente em nenhum deles, no entanto, os valores sustentam a conclusão de que o nosso YouTube se aproxima mais dos perfis 0-19.

```
(0-1) (0.023333333333333317, 0.99999673514201559)
(0-30) (0.13, 0.011285883215158372)
(0-youtube) (0.11256157635467978, 0.023180276109979574)
(30-youtube) (0.15589490968801312, 0.00038694974170490543)
```

Figura 12: Resultado do teste Kolmogorov–Smirnov (KS) para os vários perfis

9 Exercício 8

No exercício 8 era pedido para inferir as pdf's dos serviços mas desta vez considerando 2 variáveis, download e upload. Os gráficos aqui apresentados estão a representar probabilidades através de cores, onde a cor azul representa uma probabilidade baixa e uma cor vermelha representa uma probabilidade alta.

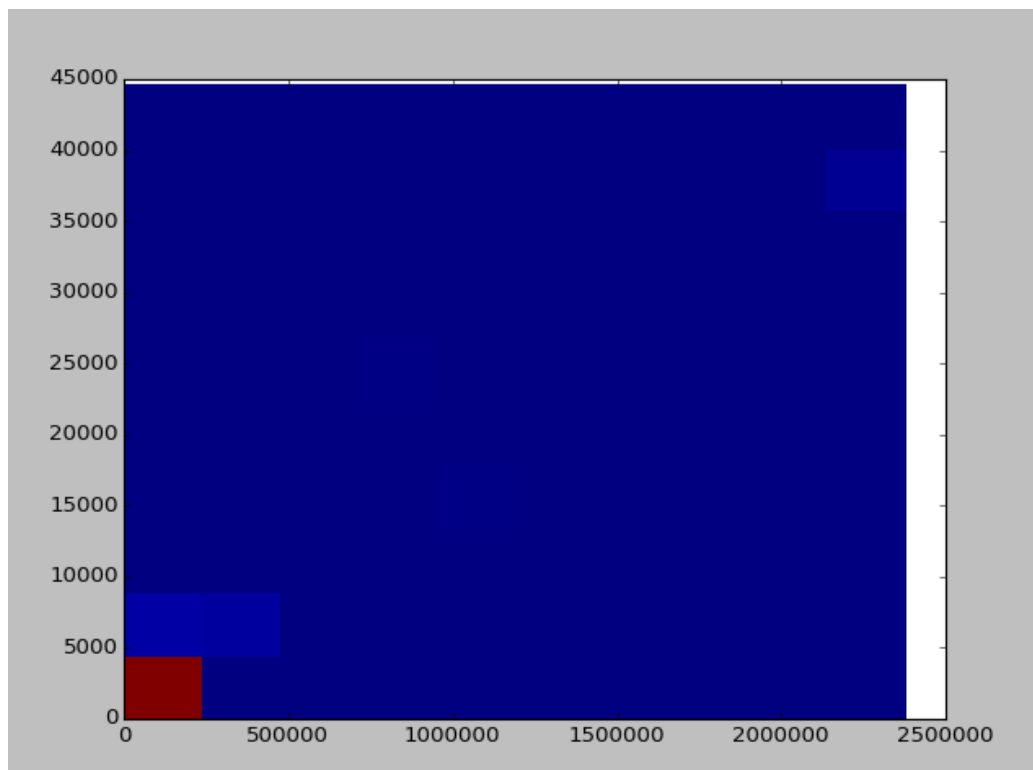


Figura 13: PDF conjunta de Download/Upload para o serviço 0

Olhando para o gráfico da figura 13 que corresponde ao serviço 0, ou seja do primeiro grupo de serviços, podemos concluir, pelo quadrado vermelho, que a maioria dos pacotes se encontram perto do 0 para o download e upload. Este fato é consistente com o que tem vindo a ser observado ate agora em que como os pacotes são recebidos em bursts isto faz com que na maior parte do tempo não sejam recebidos quaisquer pacotes. Se olharmos com mais atenção é possível ver uns quadrados a azul mais claro que representam os bursts onde são recebidos grandes quantidades de pacotes. Podemos ainda ver os valores máximos de largura de banda que foi exigido de download e upload pelo serviço correspondendo estes a 2.5Mb e 45kb respetivamente.

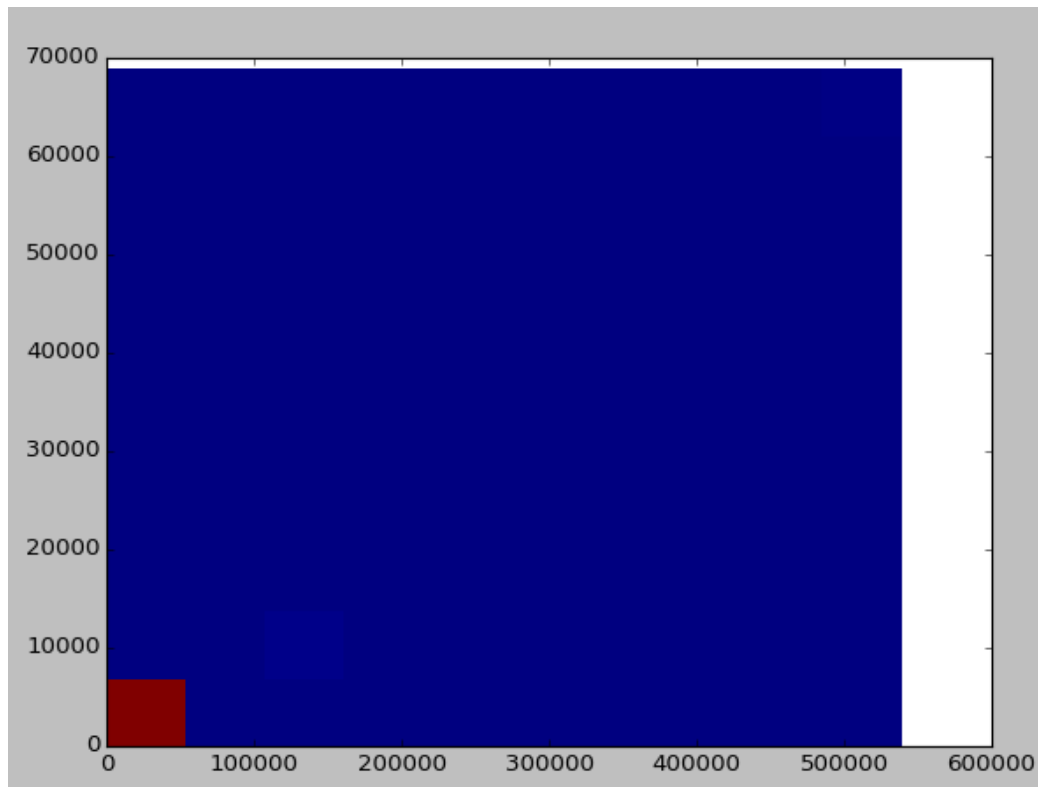


Figura 14: PDF conjunta de Download/Upload para o serviço 20

Quando olhamos agora para a figura 14 é possível encontrar semelhanças entre este gráfico e o anterior. Ambos os serviços tem uma maior probabilidade de receber pacotes perto de zero, mais uma vez representado pela cor vermelha. Uma primeira diferença que é possível identificar é a não existência dos quadrados mais claros em valores tão elevados como no gráfico da figura 13. Neste caso existe apenas um quadrado com probabilidade ainda mais pequena, comparativamente ao gráfico anterior, de existirem pacotes com maior quantidade. Analisando com mais pormenor o gráfico é possível identificar uma outra diferença na quantidade de tráfego produzida por ambos os serviços. Este serviço produziu no máximo um pico de 600kb e 70kb de download e upload respetivamente.

Comparando os valores de pico de largura de banda para os dois serviços podemos dizer que se encontram dentro do esperado. O primeiro serviço, de Youtube, atingiu um pico de 2.5Mb/s de download e em contrapartida usou de upload apenas 45kb/s uma vez que para este serviço apenas acknowledge's são necessários enviar para os servidores. Olhando para o segundo serviço, de navegação na web, vimos que atingiu picos de 600kb/s e 70kb/s de download

e upload respetivamente. Estes valores são também dentro do esperado, visto que apenas esta a carregar páginas web daí o baixo download em relação ao primeiro serviço. Ao nível do upload seria esperado um valor mais elevado do que no primeiro serviço uma vez que além dos acknowledge's são também enviados pedidos aos servidores e alguns dados que o utilizador insira nas páginas web.

10 Exercício 9

No exercício 9 era pedido para inferir as pdf's dos agregados de vários serviços(20 a 500), e comparar com uma distribuição Normal/Gaussiana com a mesma média e variância. Apresentamos de seguida os vários gráficos resultantes de diferentes números de agregações.

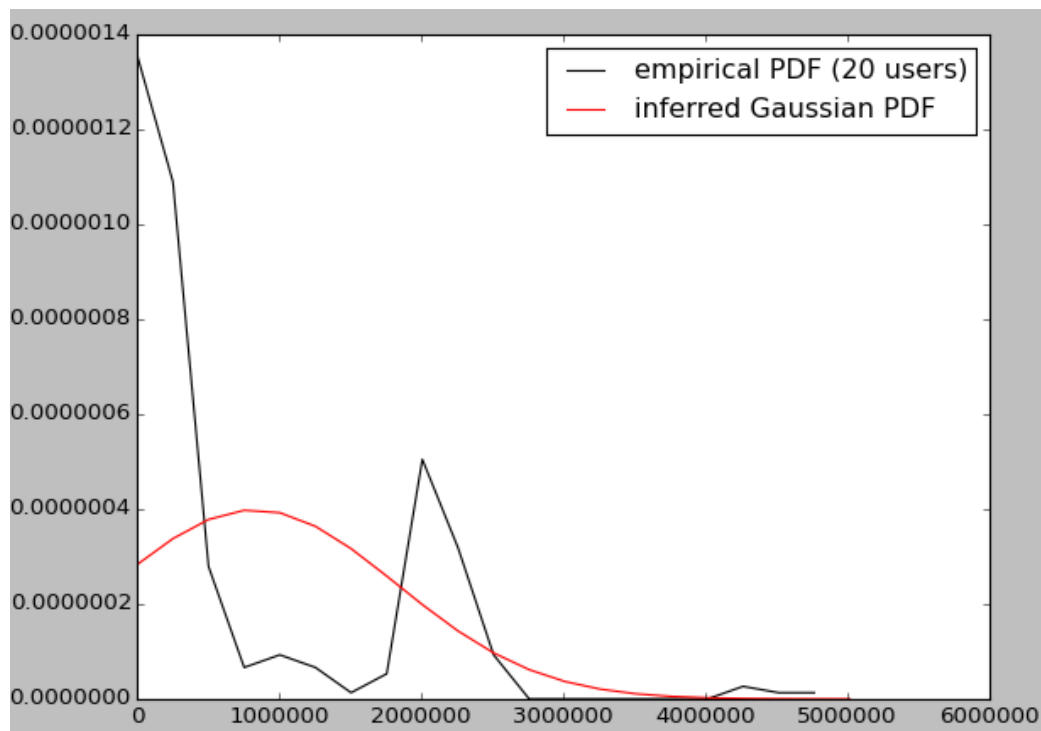


Figura 15: PDF do agregado de 20 serviços

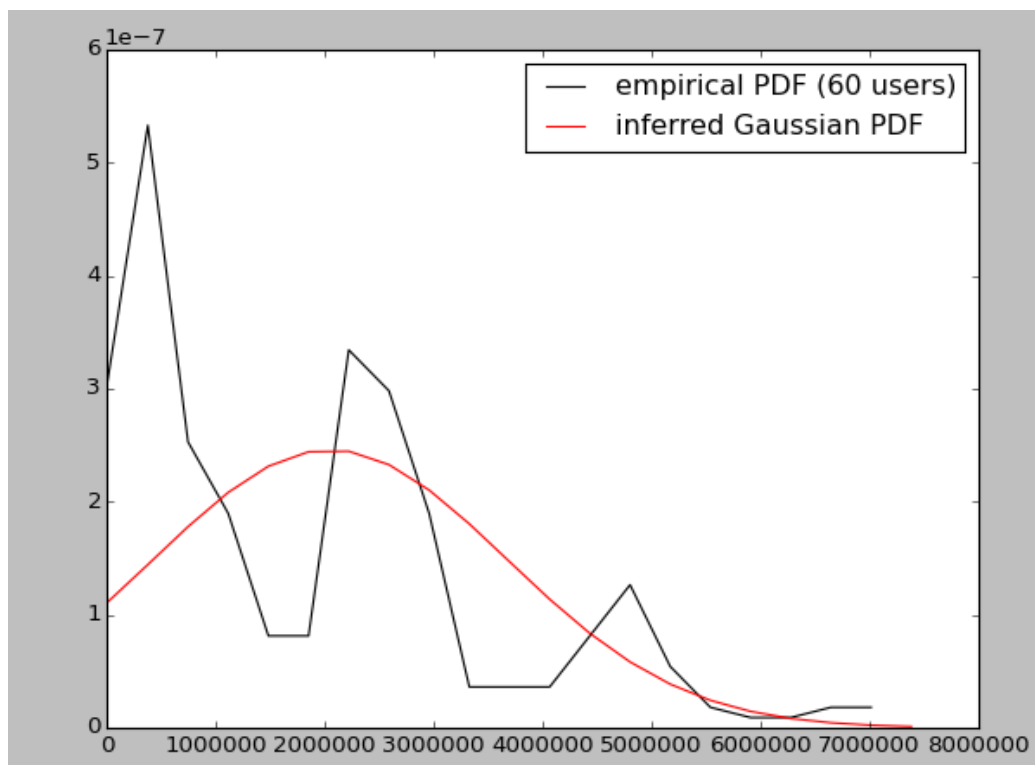


Figura 16: PDF do agregado de 60 serviços

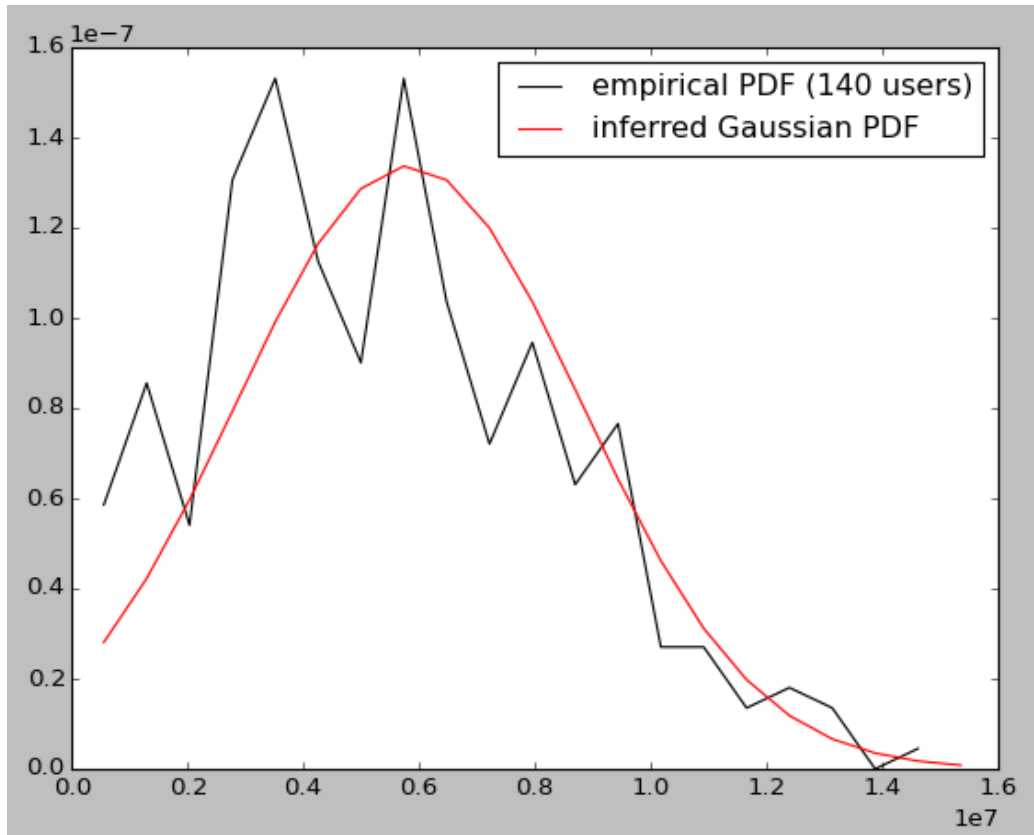


Figura 17: PDF do agregado de 140 serviços

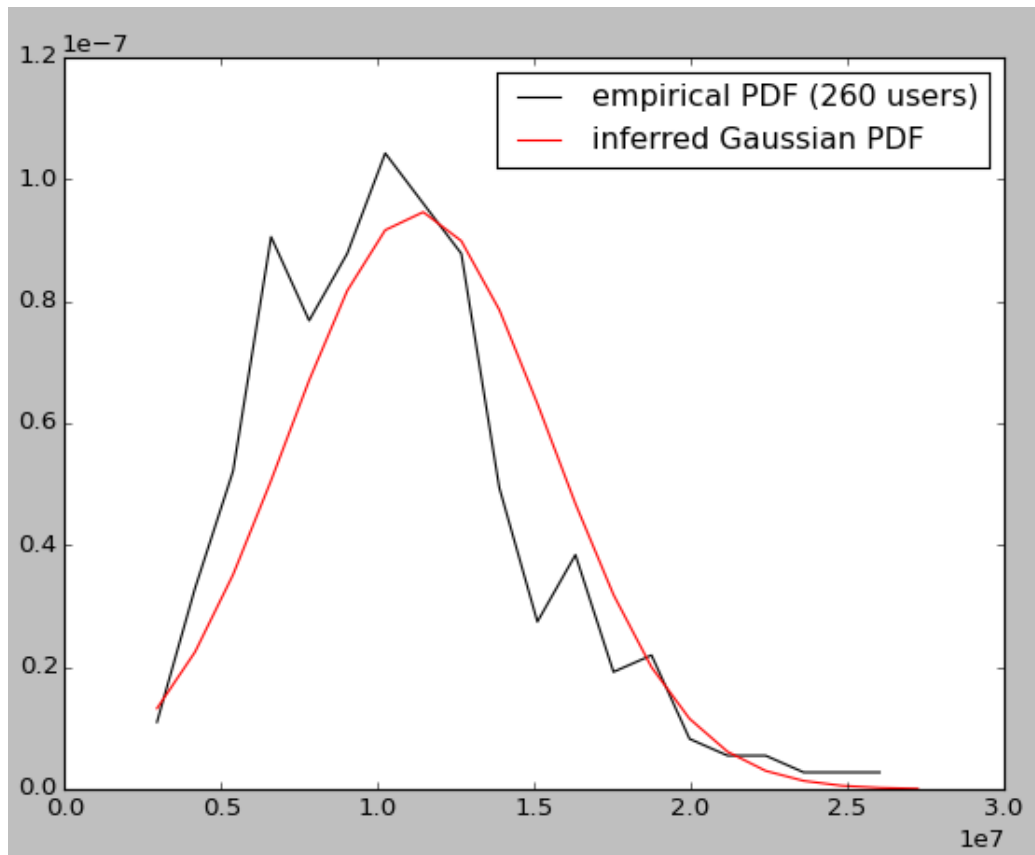


Figura 18: PDF do agregado de 260 serviços

Como é possível verificar, para 20 serviços agregados, não há grande semelhança com a pdf gaussiana gerada, no entanto, aumentando o número de agregações, é observável que a PDF dos serviços se começa a aproximar cada vez mais da curva gaussiana.

11 Exercício 10

No exercício 10 era dado um ficheiro que continha os dados de bytes de tráfego para 10 links numa rede. Era ainda dito que foi detetada uma anomalia no primeiro link (link 1) e era pedido para identificar baseado numa matriz de correlação dos dados entre os links quais seriam os restantes links afetados. Essa matriz está representada na figura em baixo como um gráfico onde a cor vermelha representa uma correlação mais forte e uma cor azul escura representa que não existe correlação.

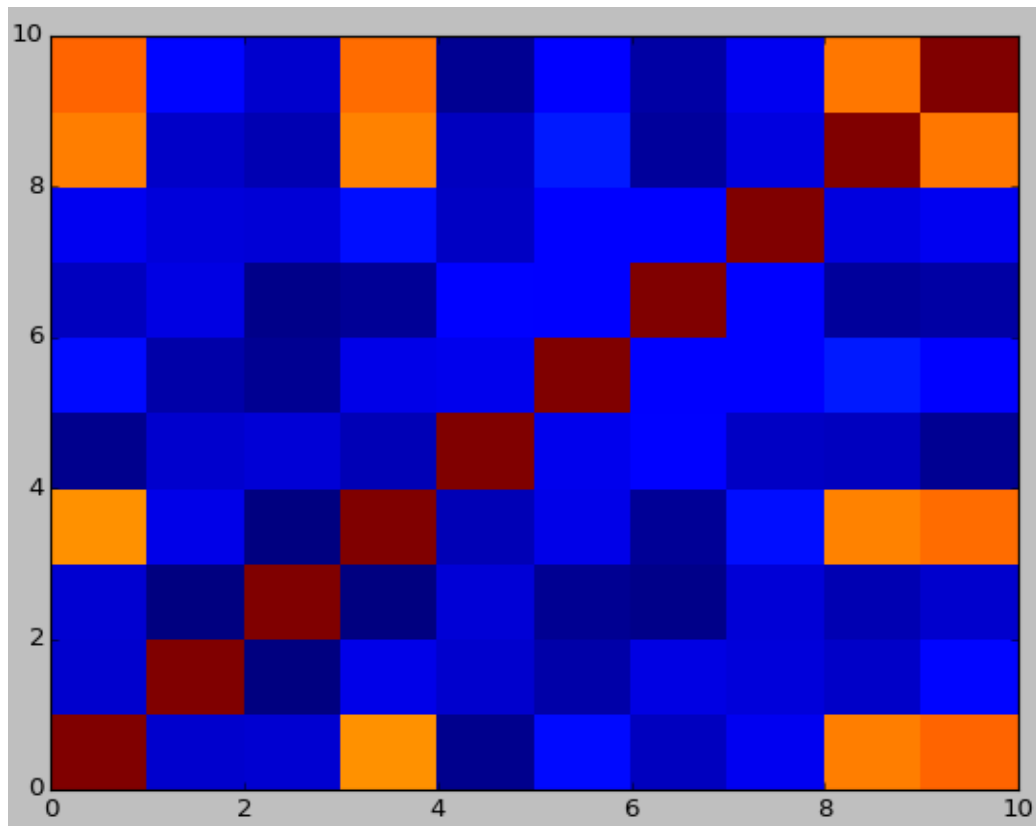


Figura 19: Matriz de correlação entre os 10 links

Analisando a matriz de correlação dos 10 links a primeira coisa a identificar é que a correlação de cada link consigo próprio tem o valor de 1 daí a cor vermelha de correlação máxima. Olhando agora para a primeira coluna e sabendo que existiu uma anomalia no link 1, é possível concluir que essa anomalia afetou o link 4, 9 e 10 uma vez que existe uma maior correlação com esses links. Isso é visível na matriz pelas cores mais laranjas em contraste às cores azuis nos restantes links.

12 Exercício 11

No exercício 11 era pedido para inferir a autocorrelação entre serviços dos dois grupos de tráfego e analisar e comprar a periodicidade entre ambos.

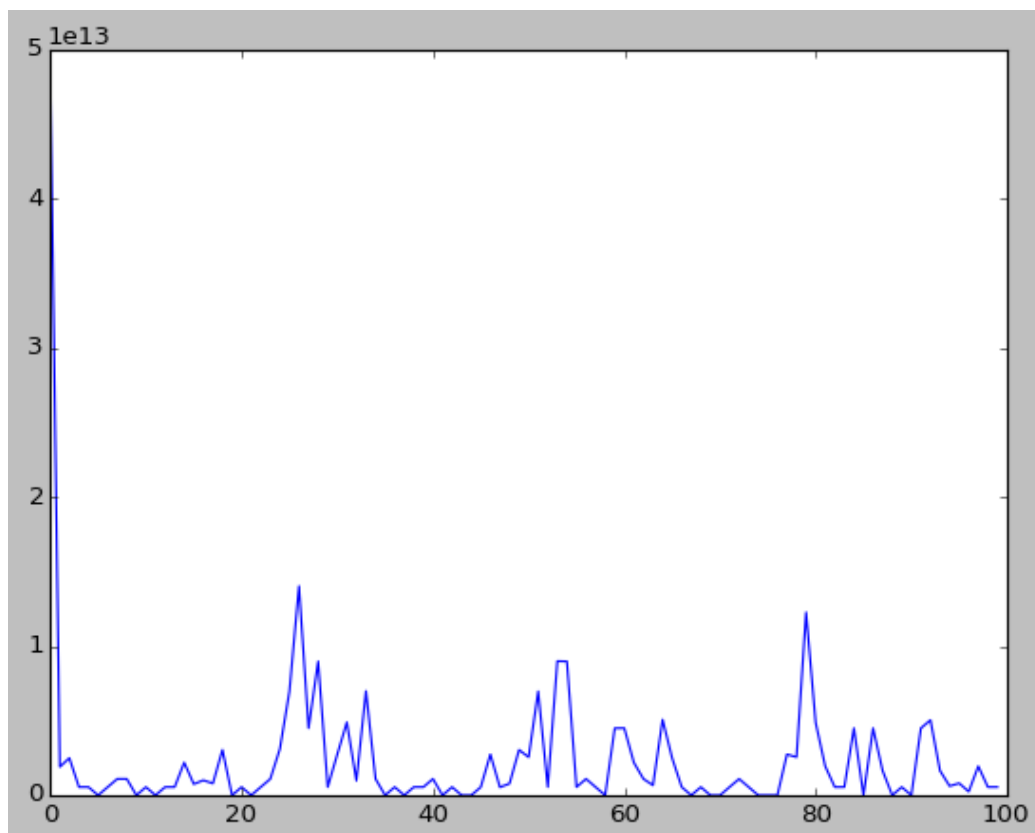


Figura 20: Autocorrelação do serviço 1

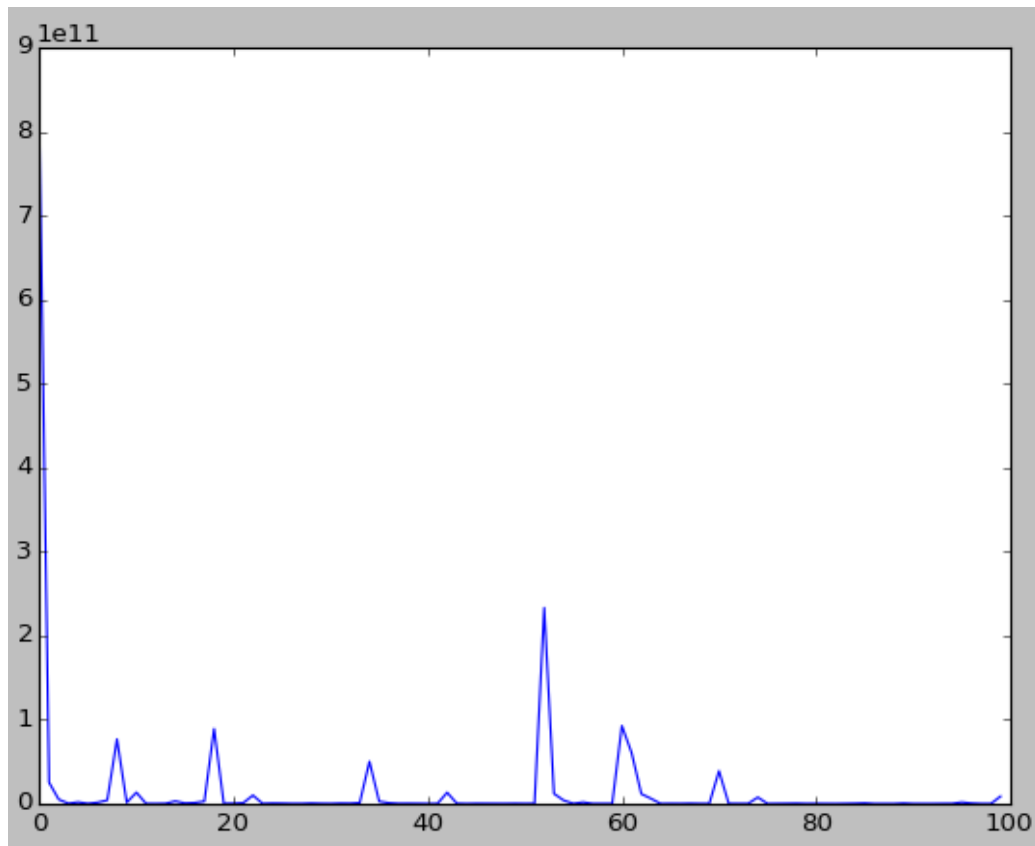


Figura 21: Autocorrelação do serviço 20

Quando olhamos para a figura 20 podemos ver que os picos tem um período aproximadamente constante. Isto seria de esperar, pois como sabemos, o serviço do Youtube funciona por envio de pacotes em bursts constantes ao longo do tempo.

Observando agora a figura 21 podemos ver que os picos não são tão periódicos como na figura 20. Este é um resultado esperado uma vez que sendo este um serviço de navegação na web, este esta dependente da interação do utilizador.

13 Exercício 12

No exercício 12 era pedido para inferir o periodograma para alguns perfis de tráfego de download usando o método de Welch.

Olhando para os gráficos a baixo é possível identificar que existe periodicidade em ambos os perfis de tráfego. Embora no primeiro perfil os picos

tenham amplitudes semelhantes e no segundo perfil os picos são mais irregulares, o próprio período não é tão constante.

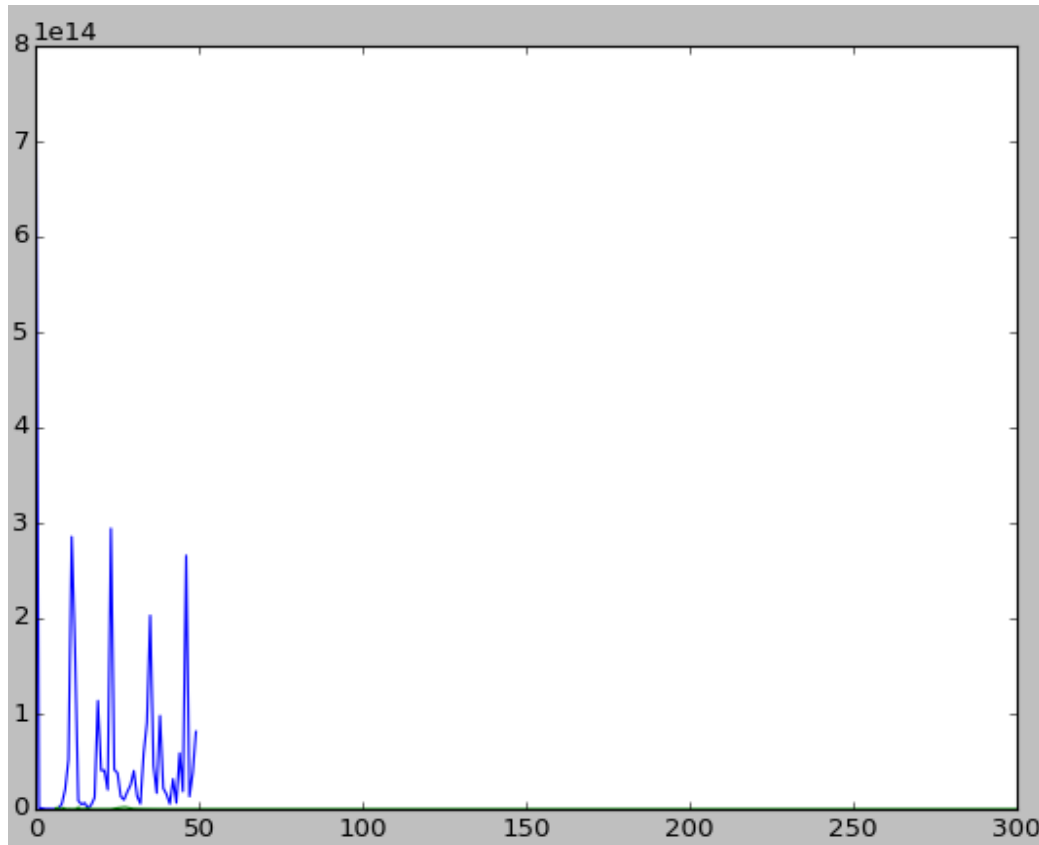


Figura 22: Periodograma para o perfil de tráfego 1

Quando olhamos para a figura 22 podemos ver que os picos tem uma amplitude aproximadamente constante e um período também aproximadamente constante. Mais uma vez, isto é consistente com o esperado uma vez que sabemos que o serviço do Youtube funciona por envio de pacotes em bursts constantes ao longo do tempo.

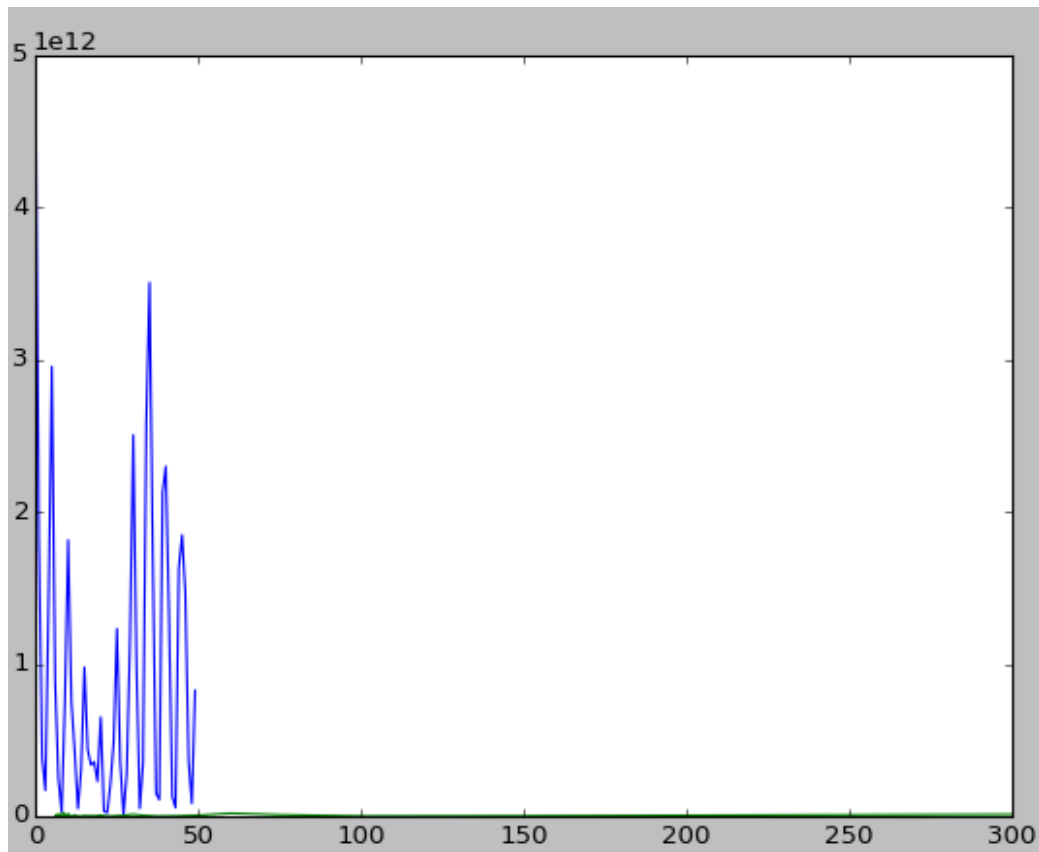


Figura 23: Periodograma para o perfil de tráfego 20

Quando olhamos para a figura 23 podemos ver que os picos tem uma amplitude irregular e embora o período tenha alguma regularidade não se assemelha ao perfil visto anteriormente. Mias uma vez, isto é consistente com o esperado uma vez que sabemos que o segundo serviço, de navegação web, está dependente da interação do utilizador para a receção de pacotes.

14 Exercício 13

No exercício 13 era pedido para inferir o escalograma para alguns perfis de tráfego, com um CWT usando o algoritmo de FFT, e usando a Morlet wavelet.

Nas figuras seguintes, apresentamos os gráficos resultantes para serviços de diferentes grupos.

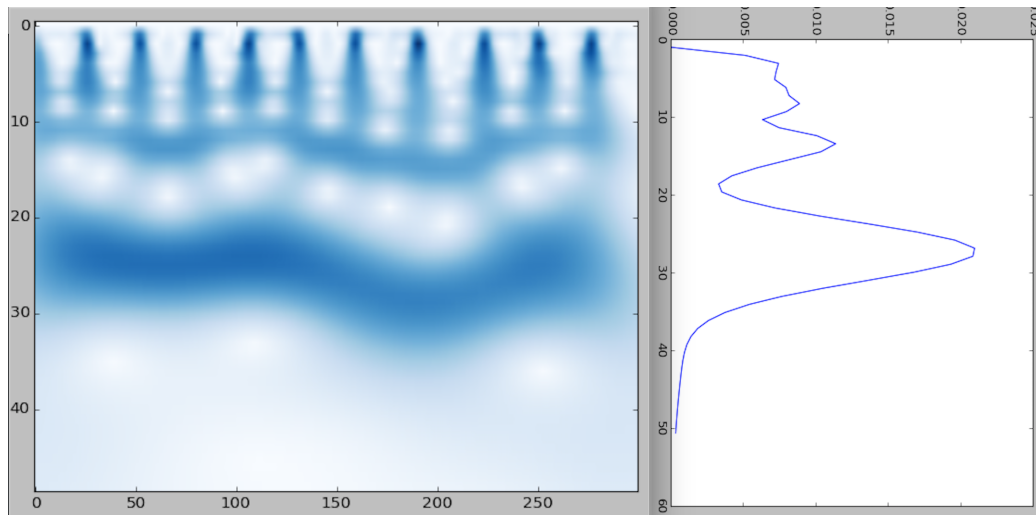


Figura 24: Escalograma para o perfil de tráfego 20

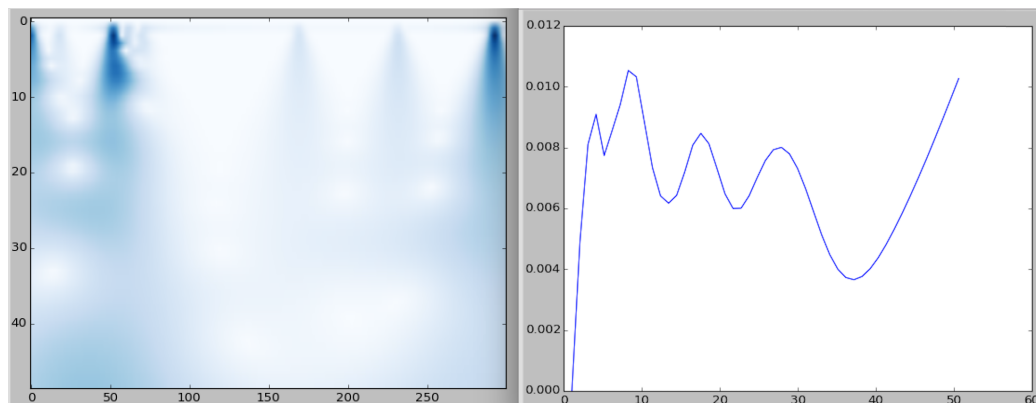


Figura 25: Escalograma para o perfil de tráfego 20

Como é possível observar nos escalogramas, há zonas com várias intensidades de azul. Rodando o gráfico ao lado do escalograma, podemos verificar que as zonas mais escuras correspondem a picos de energia(frequência) mais elevados. Enquanto que as zonas mais claras do escalograma correspondem a zonas com menos amplitudes.

15 Exercício 14

No exercício 14 era pedido para efetuar a redução de variáveis por forma a conseguir identificar e diferenciar diferentes serviços usando os perfis de tráfego que têm sido estudados. Esta redução de variáveis foi realizada usando a análise de componentes principais que usa uma matriz de covariância

de forma a identificar que componentes usar na comparação entre os diferentes componentes. Não é possível saber que dois componentes foram usados.

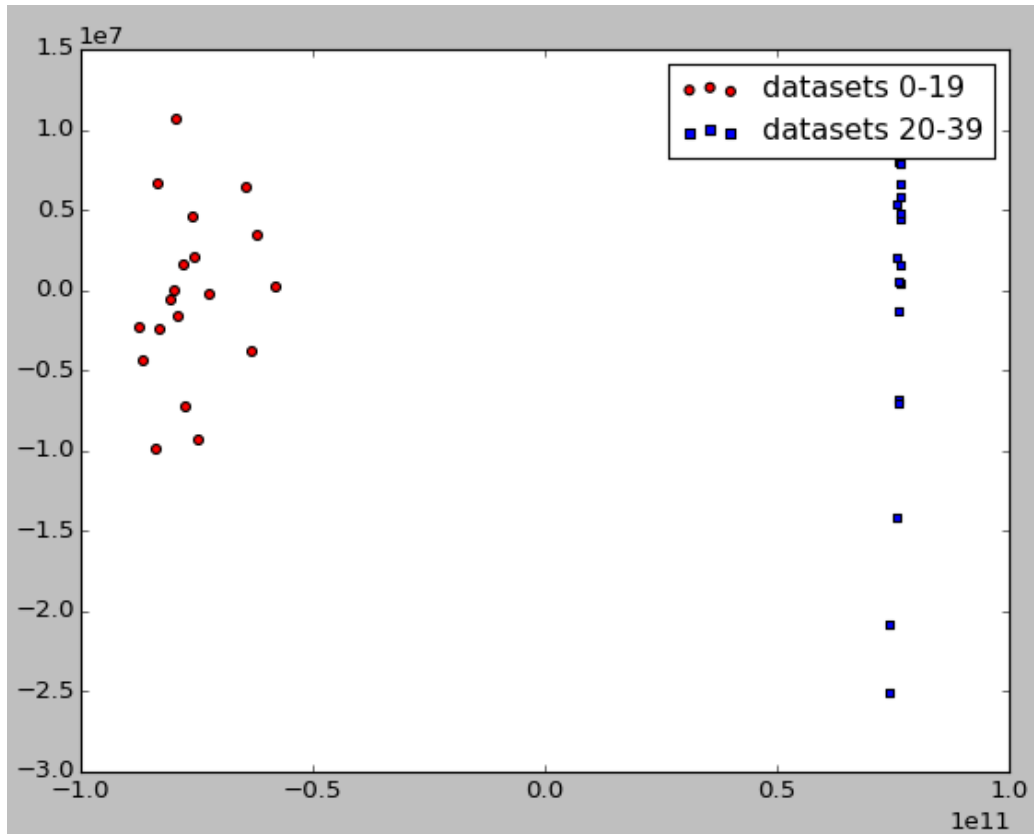


Figura 26: Representação dos perfis de tráfego após a redução de variáveis

Olhando agora para a figura 26 podemos facilmente distinguir dois serviços que vem a sustentar as nossas conclusões até ao momento.

Os perfis de tráfego correspondentes ao primeiro serviço (Youtube) correspondem aos pontos vermelhos na figura e encontram-se distribuídos em duas componentes embora que numa área não muito distante uns dos outros.

Os pontos azuis representam os perfis de tráfego correspondentes ao segundo serviço (Navegação Web) e encontram-se bastante afastados dos perfis do serviço anterior, ou seja dos pontos vermelhos. Podemos ver ainda que estes perfis variam apenas em uma componente ao contrario dos anteriores que variavam em duas.

16 Exercício 15

No exercício 15 era pedido para fazer uma análise de agrupamentos de serviços usando o algoritmo K-mean. Este algoritmo usa a distancia entre pontos como métrica e requer que seja dito à priori o número de agrupamentos pretendido.

Nas figuras em baixo estão representados os gráficos relativos à aplicação do algoritmo referido com 2, 3 e 4 agrupamentos.

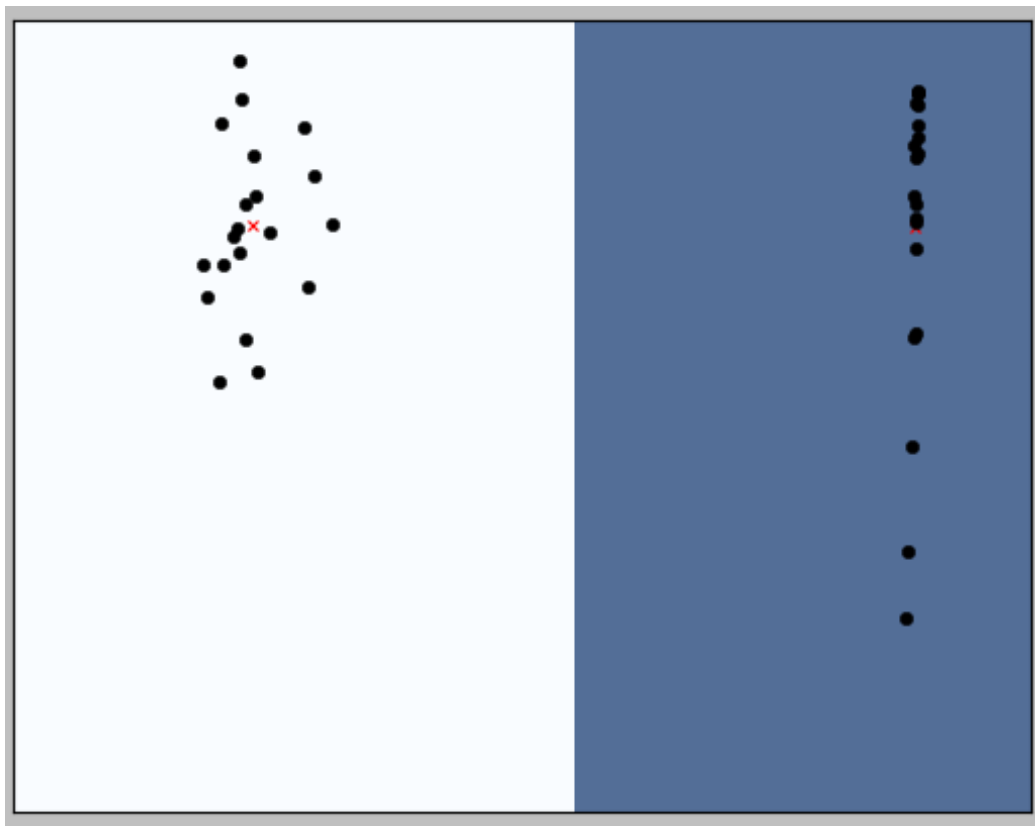


Figura 27: K-mean com 2 agrupamentos

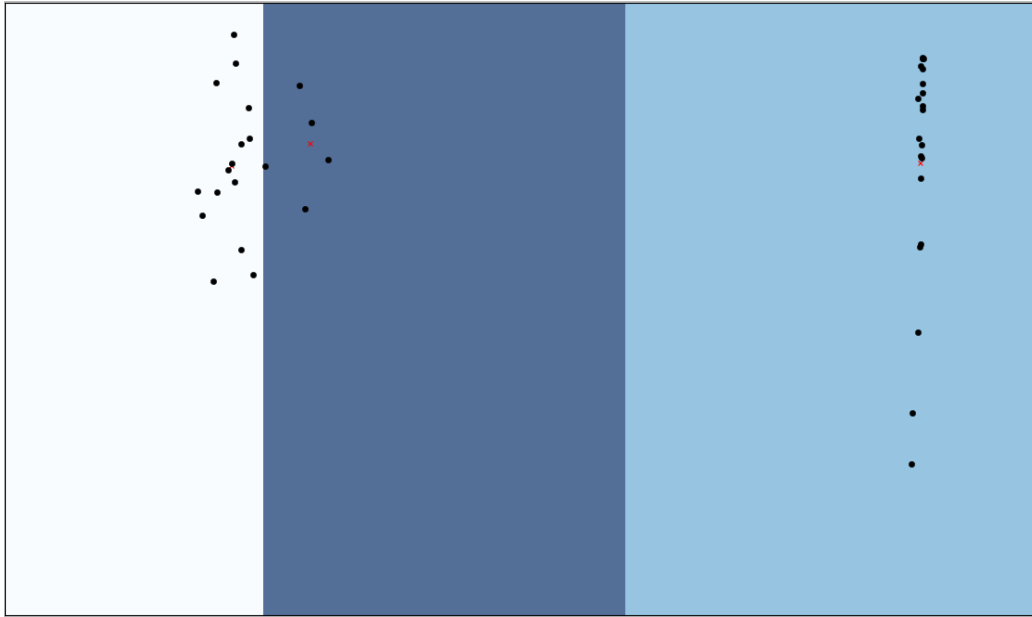


Figura 28: K-mean com 3 agrupamentos

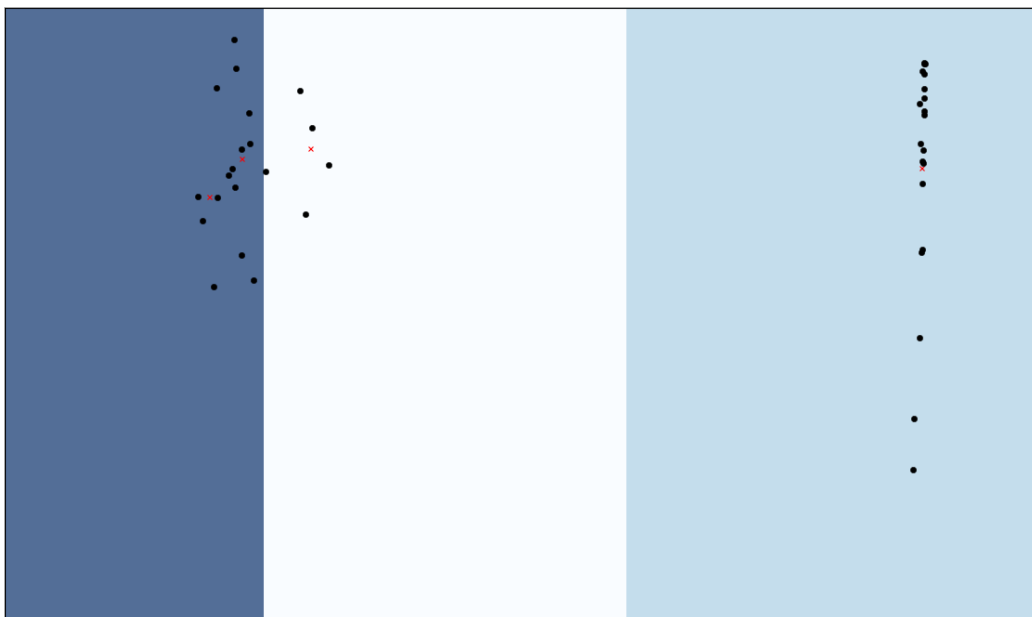


Figura 29: K-mean com 4 agrupamentos

Observando a figura 27, em que foram usados dois agrupamentos, nota-se que houve uma divisão clara entre os dois grupos de serviços. Na figura 28, com 3 agrupamentos, o primeiro grupo de pontos foi dividido em dois,

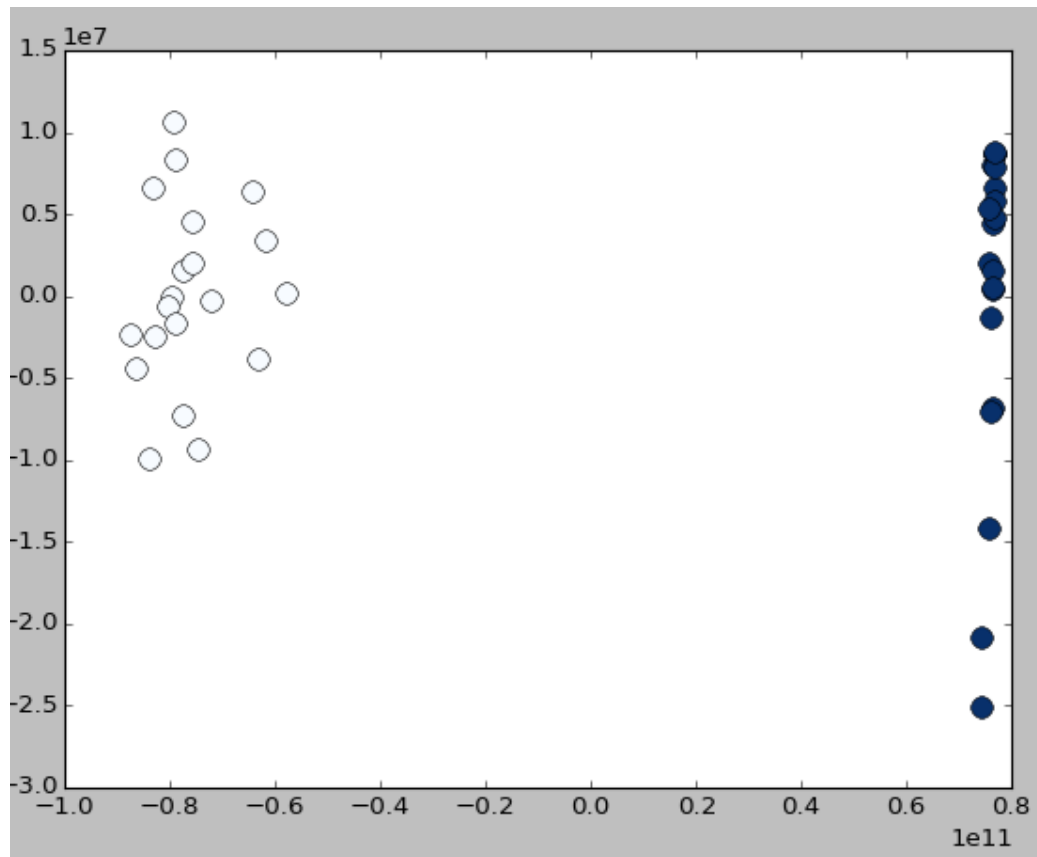
talvez devido a estes se encontrarem mais dispersos nas duas componentes do que os pontos do outro grupo. Esta divisão pode levar a que se interprete mais um serviço que na realidade não existe. Por fim, na figura 29, com 4 agrupamentos, o primeiro grupo ficou ainda mais dividido.

Analisando estes resultados, é seguro afirmar que sendo o k-mean um algoritmo que requer à priori o número de agrupamentos, se houver uma introdução errada deste número, o k-mean vai inferir uma divisão que na realidade não existe.

17 Exercício 16

No exercício 16 era pedido para fazer agrupamento de serviços usando o algoritmo DBSCAN. Este algoritmo usa a distancia entre os pontos mais perto como métrica e requer que seja dito à priori o tamanho da vizinhança pretendido.

Em baixo estão representados os gráficos relativos à aplicação do DBSCAN com vizinhanças de $1e10$ e $1e11$ e $1e8$ respetivamente.

Figura 30: DBSCAN com vizinhança de $1e10$

Olhando para o gráfico da figura 30 produzido com vizinhança de $1e10$ é possível identificar que a divisão dos grupos foi realizada conforme temos vindo a concluir ao longo deste relatório, havendo dois grupos que correspondem aos dois serviços, Youtube e Navegação Web.

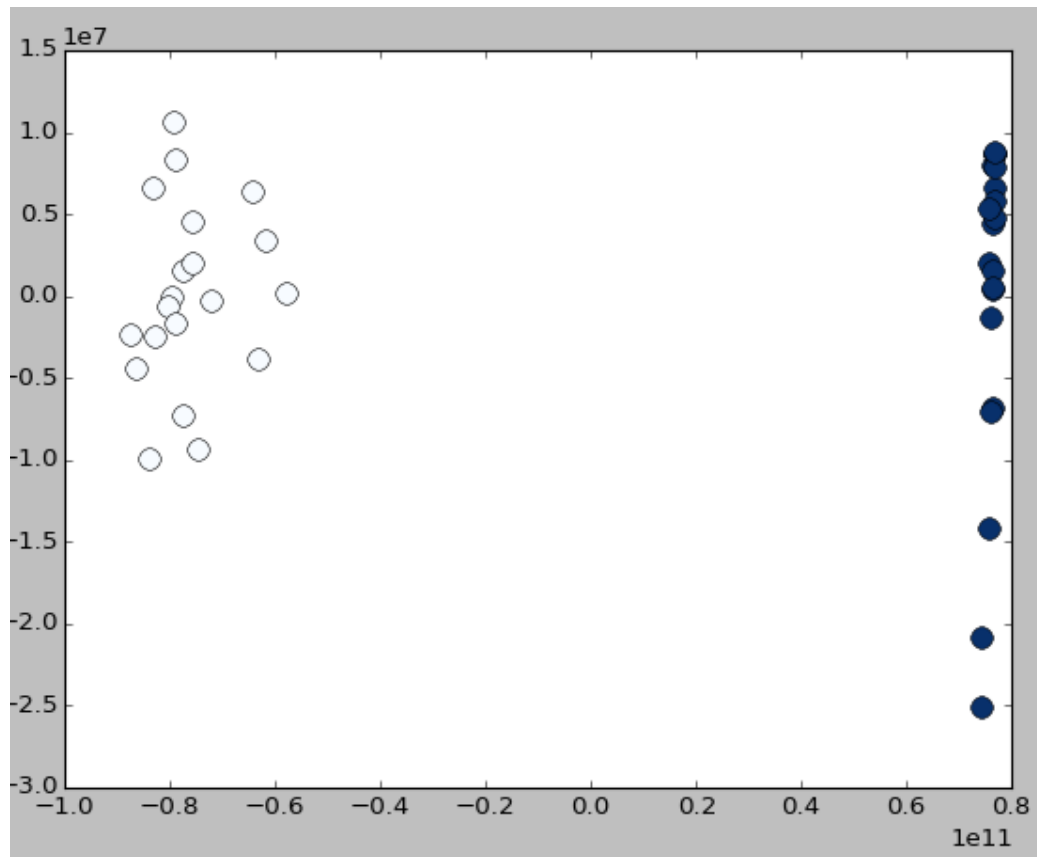
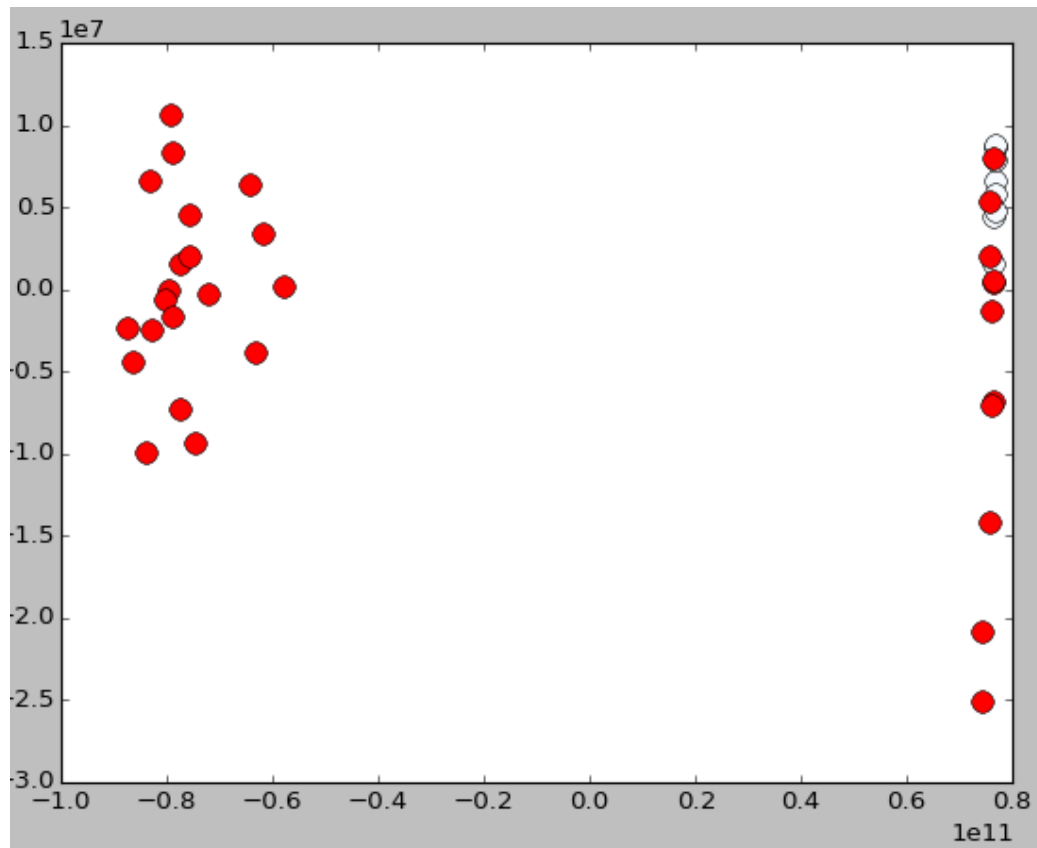


Figura 31: DBSCAN com vizinhança de $1e11$

Quando olhamos para a figura 31 produzido com vizinhança de $1e11$ podemos tirar as mesmas conclusões que na figura anterior. A alteração da vizinhança não produziu diferença no aglomerado produzido.

Figura 32: DBSCAN com vizinhança de $1e8$

Quando olhamos para a figura 32 produzido com vizinhança de $1e8$ podemos ver que os resultados produzidos são diferentes do que com as vizinhanças anteriores. Com esta vizinhança podemos ver que dentro do grupo da navegação web, grupo da direita, houve alguns pontos que foram separados do aglomerado principal, isto pode ser explicado pela navegação web ser algo que depende dos cliques do utilizador e como tal podem haver alguns perfis que não se enquadram tanto comparando com os restantes.

O que gerou dúvida ao analisar este gráfico foi o facto de os dois grupos, identificados anteriormente, estarem representados pela mesma cor, o que nos faria apontar para que ambos teriam sido identificados como um único aglomerado. Esta conclusão para nós não fez muito sentido ainda para mais uma vez que o tamanho da vizinhança foi reduzido.

Dito isto não nos foi possível concluir sobre este ultimo gráfico a existência dos dois serviços como temos vindo a concluir até agora.

18 Exercício 17

No exercício 17 era pedido que assumindo que os dados têm 20% de anomalias, se inferisse os limites de anomalias usando o método *elliptic envelope*.

Na figura seguinte apresentamos o gráfico resultante.

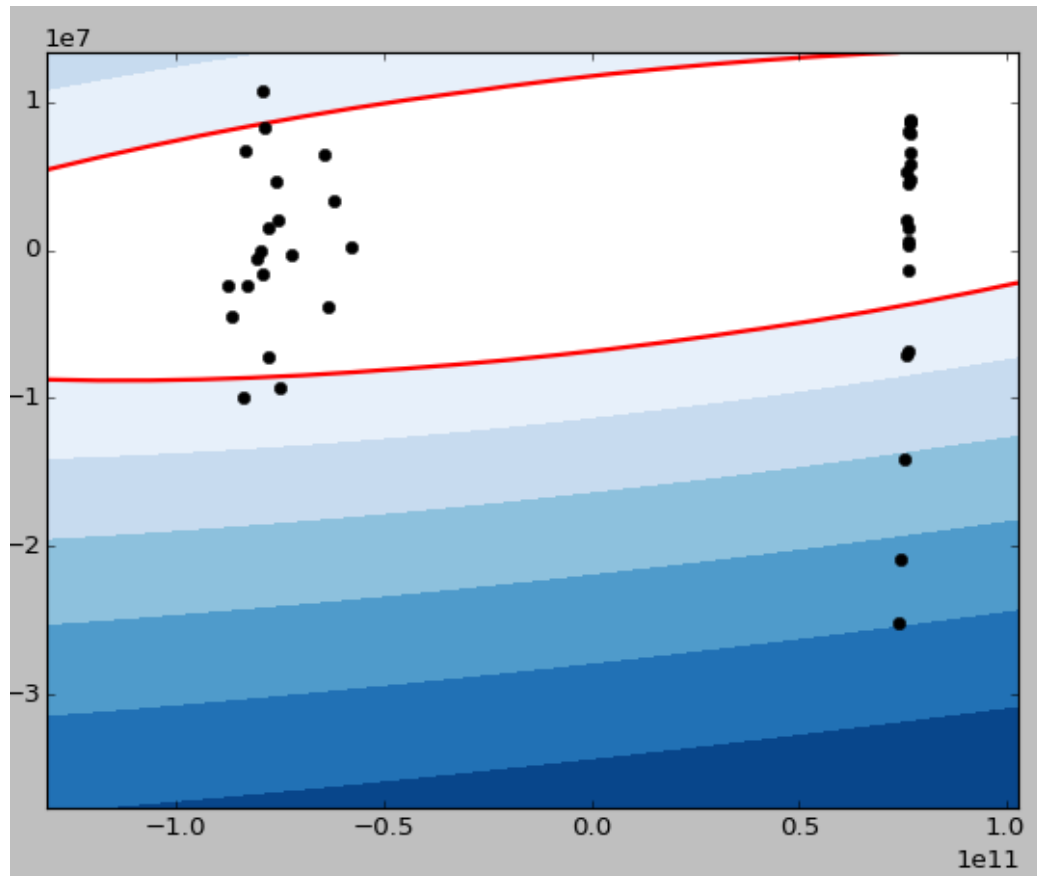


Figura 33

Observando a figura, e sabendo que as linhas vermelhas são os limites de anomalias, é possível concluir que os pontos que estão fora da zona branca são considerados anomalias. f

19 Exercício 18

No exercício 18 era pedido para gerar tráfego sintético baseado num perfil de tráfego à escolha. Foi escolhido o terceiro perfil que representa um perfil do serviço do Youtube. Usando como base a média, variância e pdf do serviço foi possível gerar os perfis de tráfego que se pode ver em baixo.

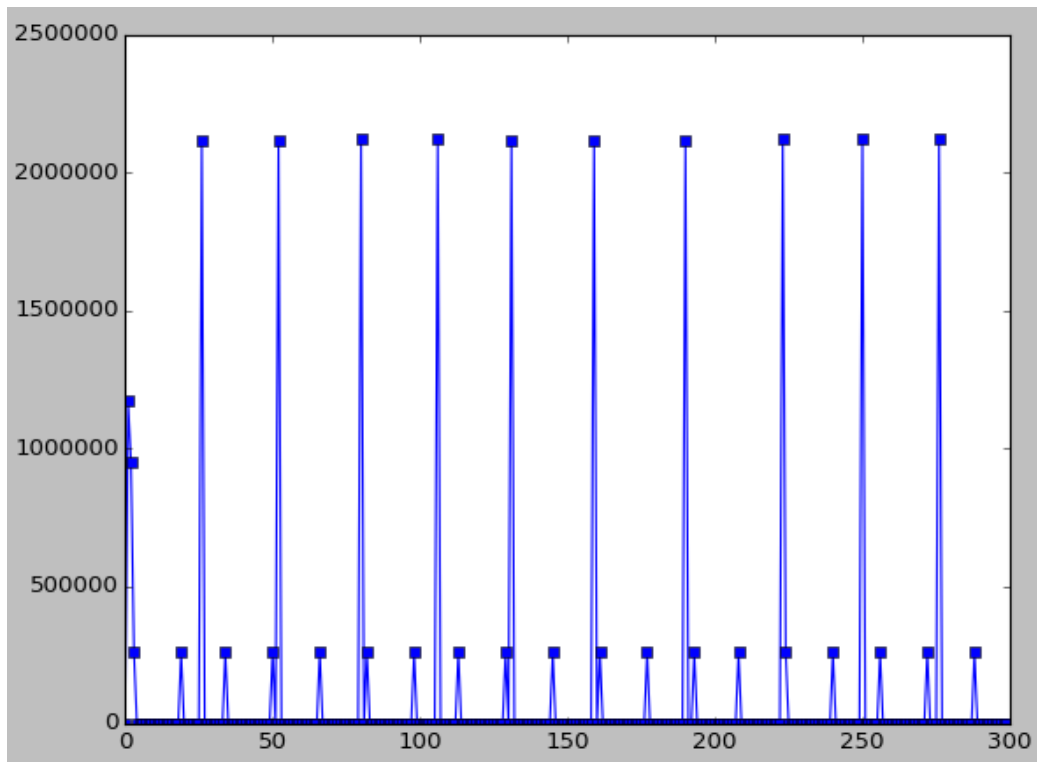


Figura 34: Representação do tráfego real

Na figura 34 temos uma representação do tráfego real que vai ser usado como comparação dos tráfegos gerados e apresentados de seguida.

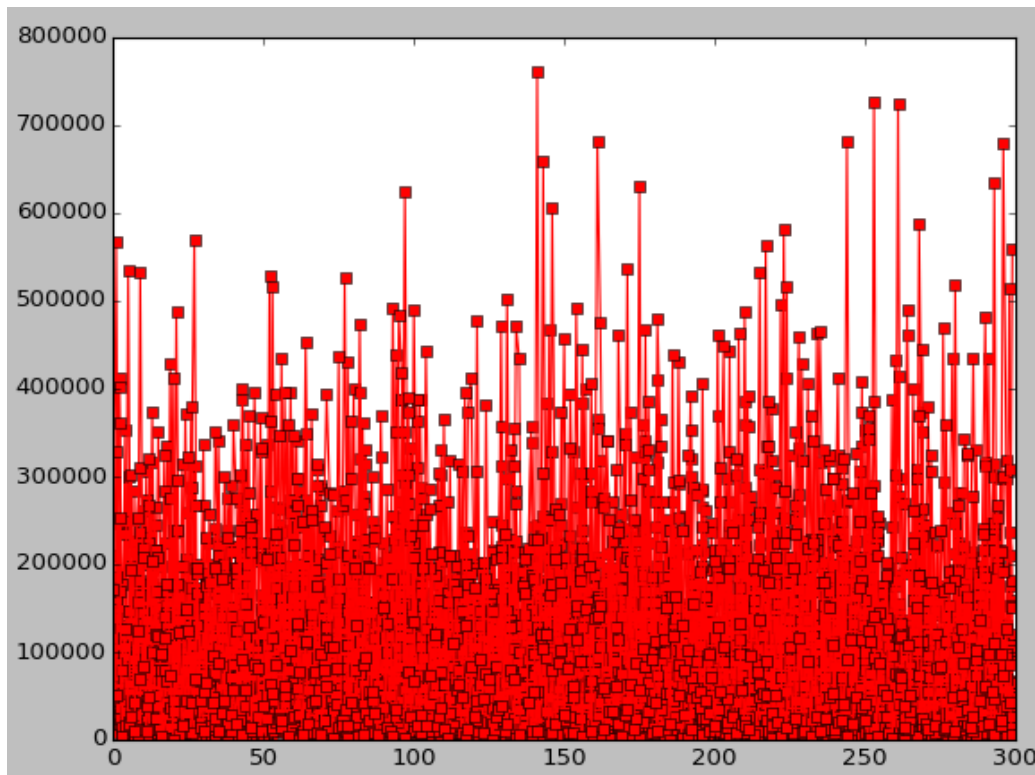


Figura 35: Representação do tráfego sintético baseado numa distribuição exponencial

Olhando para a figura 35 que corresponde ao tráfego gerado com uma distribuição exponencial podemos ver que são gerados muitos pontos que não se enquadram com o tráfego real isto é devido ao fato de que a geração deste tráfego depende apenas da média e variância do tráfego original o que faz com que se perca muita definição do tráfego original.

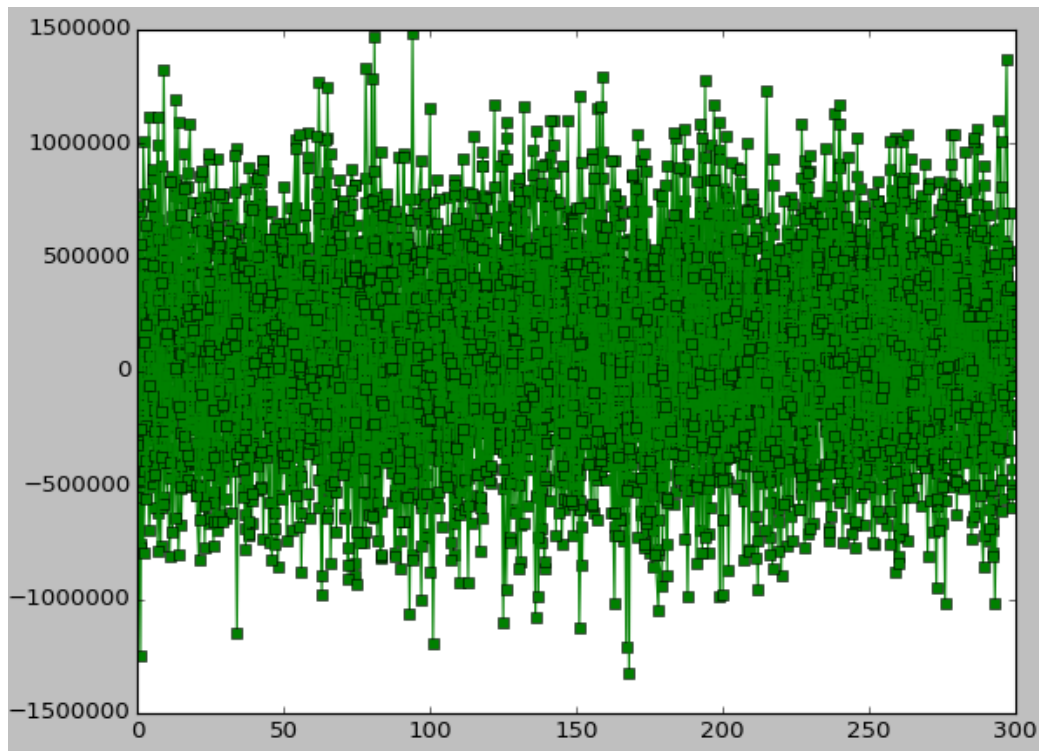


Figura 36: Representação do tráfego sintético baseado numa distribuição gaussiana

Olhando para a figura 36 temos uma representação do tráfego gaussiano que podemos ver que tem o mesmo problema do exponencial no sentido que depende apenas da média e variância e devido a este facto perde muita da definição do tráfego real. Outra observação sobre o tráfego gaussiano é a existência de valores negativos devido à própria natureza da curva gaussiana. Quando olhamos para o gráfico como valores de tráfego podemos descartar a metade negativa uma vez que nunca temos tráfego negativo na semântica usada no contexto.

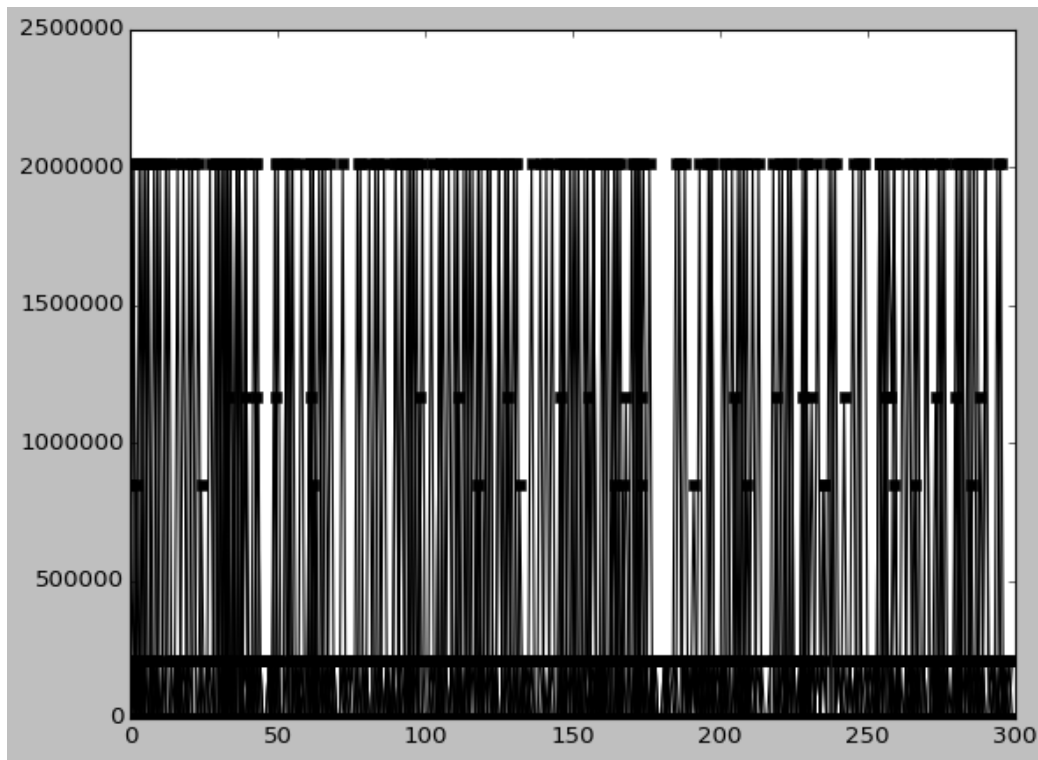


Figura 37: Representação do tráfego sintético baseado numa distribuição discreta

Olhando agora para a figura 37 que representa um tráfego sintético discreto, podemos ver que existem muito mais semelhanças com o tráfego real. Este fato deve-se pois para a geração deste tráfego é usado a pdf do tráfego real e como tal vai-se aproximar muito mais da realidade. A maior diferença deste tráfego para o real é o fato de no real existirem muitos "tempos mortos", tempos em que não é gerado qualquer tráfego, e uma vez que nesta geração de tráfego esses tempos não estão a ser tidos em consideração e esta a ser gerado um valor para cada amostragem temos um gráfico muito mais preenchido que no real. Este pormenor vai ter influencia na pdf deste tráfego como vamos ver mais a frente.

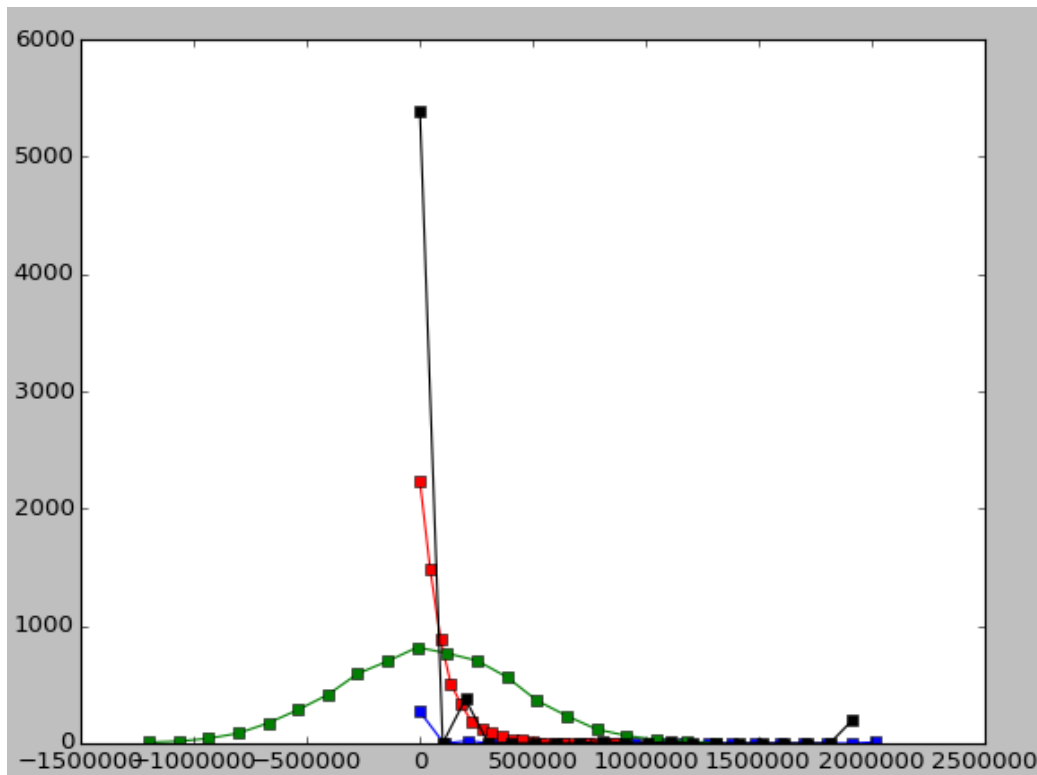


Figura 38: Comparação das pdf dos tráfegos sintéticos e real

Analisando a figura 38 podemos comparar as pdf's dos diferentes tipo de trafego com o trafego real. Nesta figura as cores das pdf representadas coincide com as cores do tipo de trafego descrito anteriormente. Para ser mais fácil a compreensão, uma vez que falta uma legenda no gráfico, temos a seguinte correspondência:

- Azul - Tráfego real
- Vermelho - Tráfego exponencial
- Verde - Tráfego gaussiano
- Preto - Tráfego discreto

Quando olhamos para as diferentes pdf's e tomando a do tráfego real como referencia podemos concluir que o tráfego gaussiano é o que menos se assemelha a este. Este tráfego além de ter a componente negativa que podemos descartar, não se assemelha em nada com o tráfego real. Olhando para a pdf do tráfego exponencial podemos ver que, embora perto do zero

esta sofra uma descida coincidente com o tráfego real, ela não tem a capacidade de representar os picos mais elevados existentes no tráfego real. Estes picos representam, como já foi dito anteriormente, os bursts de tráfego que são recebidos. Com isso, e como também já tinha sido concluído com a análise ao tráfego individual, o tráfego gerado que mais se assemelha ao real é o tráfego discreto. Esta semelhança, como também já foi dito, advém do fato que este tráfego é gerado baseado na pdf do tráfego real.

Apesar das diferenças ou semelhanças dos diferentes tipos de tráfego todos eles tem valores muito mais elevados que o tráfego original, isto deve-se ao fato que o tráfego gerado não tem em conta os "tempos mortos" que existem no tráfego real.

20 Exercício 19

No exercício 19 era pedido para escolher um perfil de tráfego e identificar os períodos de atividade(ON) e de não atividade. De seguida era pedido para modelar a duração de permanência em cada estado com uma distribuição exponencial e o tempo de atividade com uma distribuição discreta. Por fim era necessário gerar o perfil de tráfego sintético, gerar perfis ao longo do tempo e comparar a PDF e as principais estatísticas.

Para a geração do tráfego sintético foram realizados os seguintes passos:

- Reduzir o tráfego original a uns(tráfego superior a 50 Bytes) e zeros(tráfego inferior a 50 Bytes).
- Calcular a média de uns e de zeros.
- Gerar distribuições exponenciais de uns e de zeros baseados nas média calculadas anteriormente.
- Gerar a distribuição discreta do tempo de atividade com base em probabilidades calculadas a partir da PDF do tráfego original.
- Gerar o tráfego baseado em números aleatórios da distribuição de zeros e de uns para os tempos de atividade e de não atividade, preenchendo os tempos de atividade com valores da distribuição discreta gerada anteriormente.
- Cálculo da PDF do tráfego gerado e comparação com a PDF do tráfego real.

Nas seguintes figuras encontram-se os gráficos das distribuições exponenciais para os períodos de atividade e não atividade bem como a distribuição discreta dos tempos de atividade.

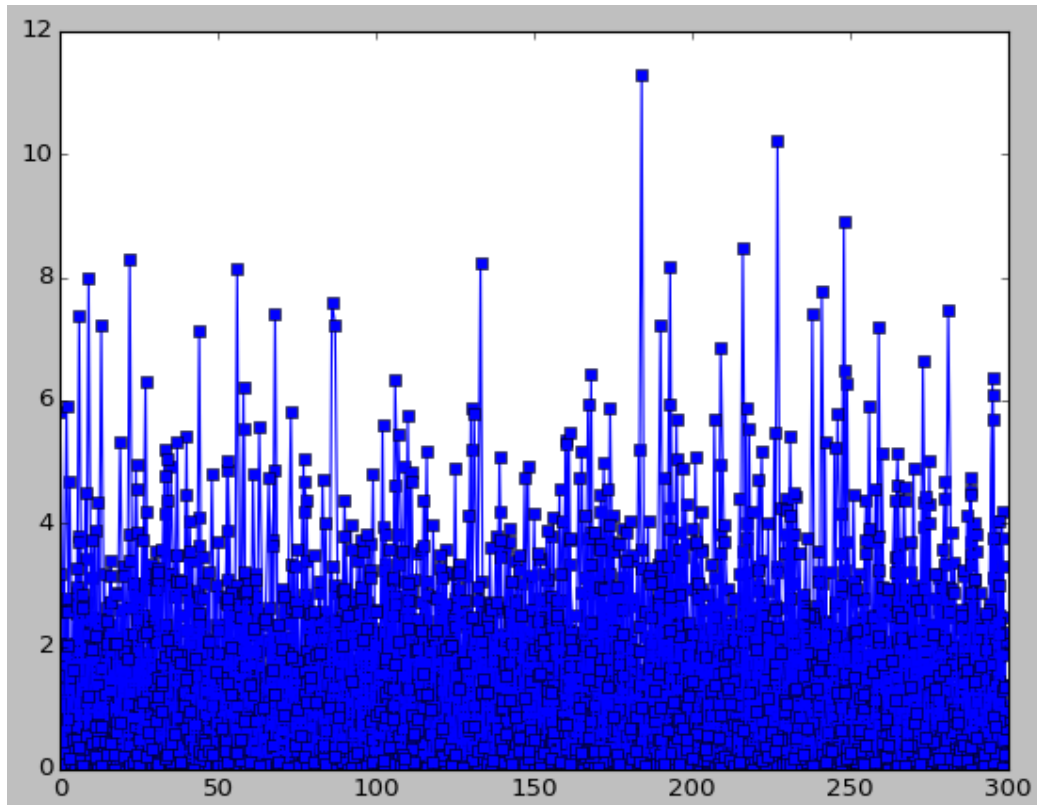


Figura 39: Distribuição exponencial em períodos de atividade

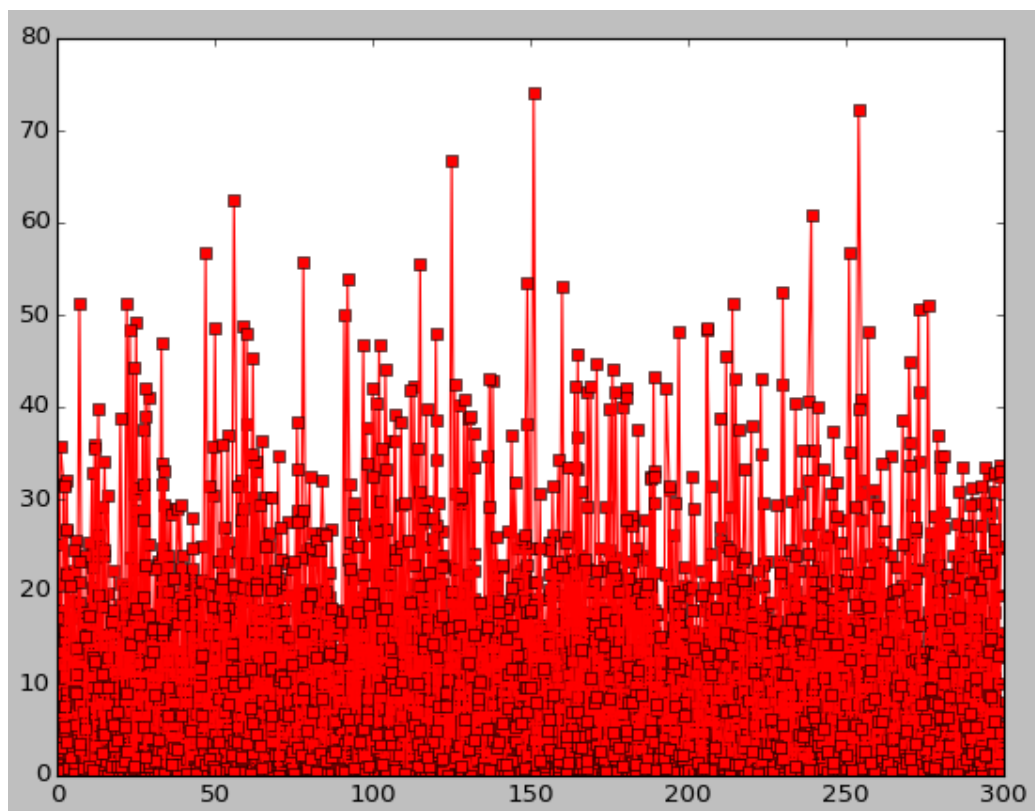


Figura 40: Distribuição exponencial em períodos de não atividade

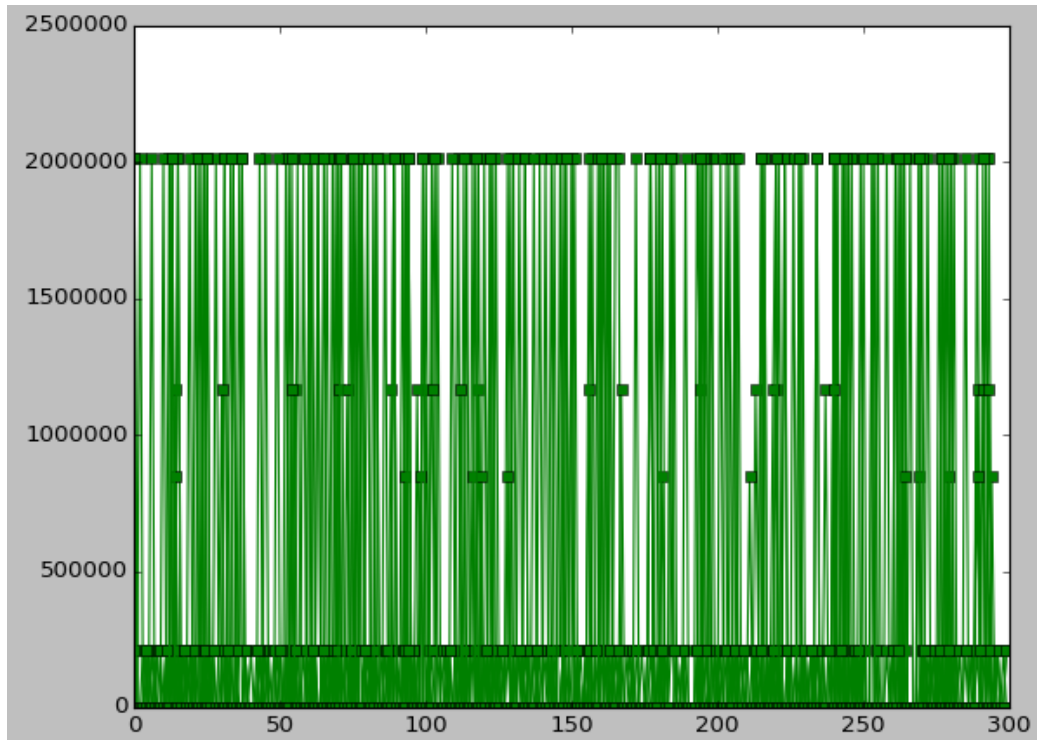


Figura 41: Distribuição discreta do tempo de atividade

Nas seguintes figuras encontra-se o gráfico do perfil de tráfego sintético, gerado a partir das distribuições calculadas anteriormente e a comparação entre as PDF's do tráfego real e deste tráfego gerado.

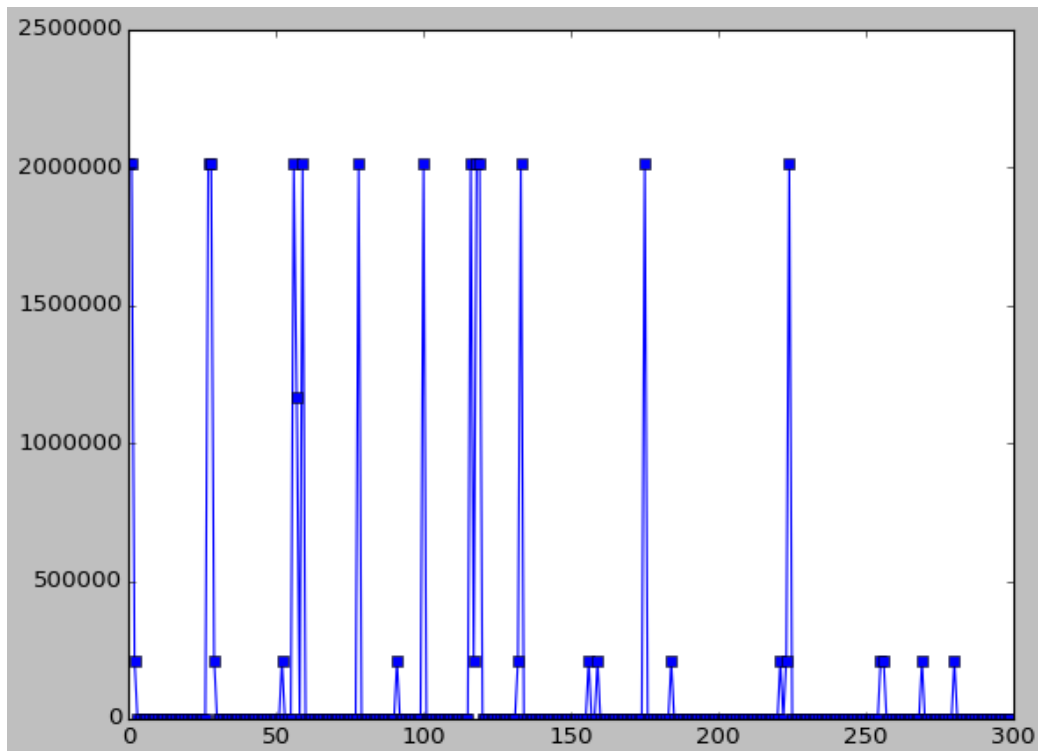


Figura 42: Perfil de tráfego sintético

Como é possível observar o tráfego gerado é bastante semelhante ao real, visto anteriormente.

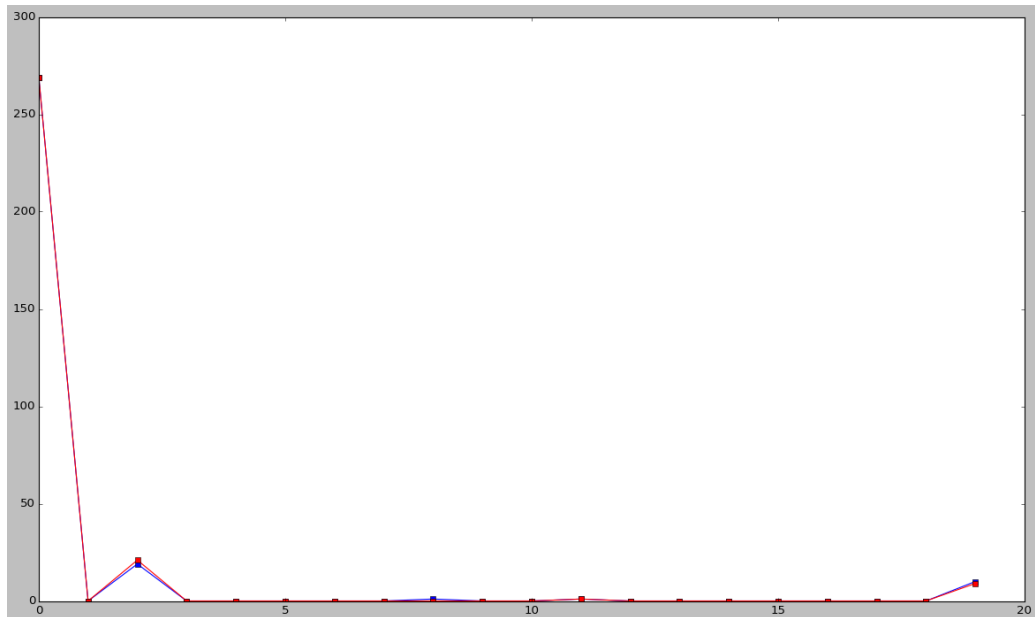


Figura 43: Comparação das PDF do tráfego real com o sintético

Analisando as PDF's, podemos reforçar a conclusão anterior sobre a semelhança dos perfis de tráfego, uma vez que as PDF's são praticamente coincidentes.

21 Exercício 20

No exercício 20 era pedido para ver qual dos modelos, linear ou exponencial, mais se enquadrava com o crescimento real que foi observado e representado aqui na figura 44 pela linha preta. Era também ainda pedido para gerar a previsão de tráfego para o link e prever quando é que atingirá os 1.5 Exabytes.

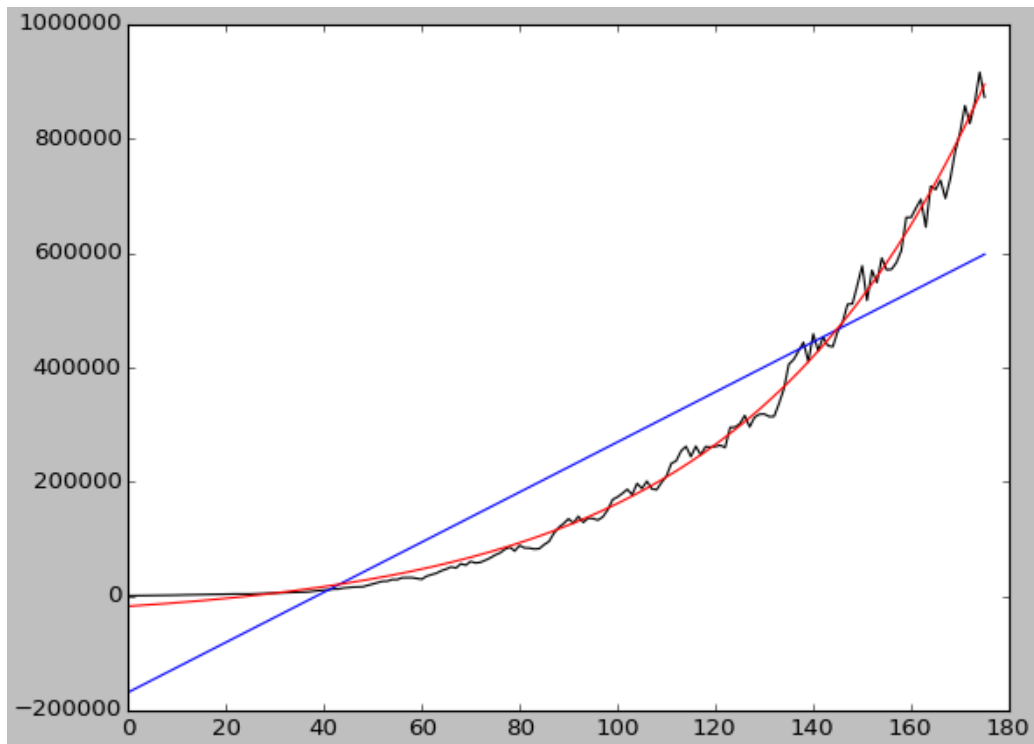


Figura 44: Aproximação do tráfego real usando um modelo linear e exponencial

Analisando a figura 44 e comparando o tráfego real (linha preta) com os modelos linear (linha azul) e exponencial (linha vermelha) é claro que o tráfego real segue um modelo de crescimento exponencial. Para conseguir estas aproximações nestes modelos foi usado um algoritmo de adaptação de curvas (curve fitting) do pacote de python "scipy".

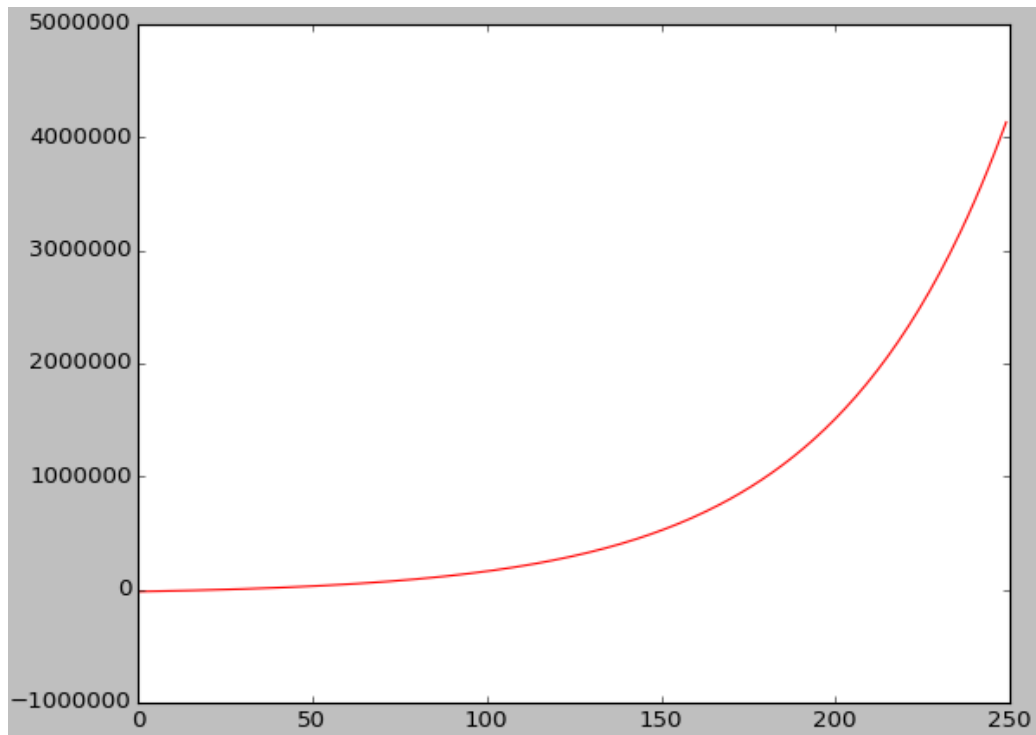


Figura 45: Previsão do tráfego no link

Na figura 45 temos a expansão da curva exponencial anterior, que foi gerada usando os parâmetros que foram obtidos através de "curve fitting", mas com valores superiores no eixo de x de forma a realizar uma previsão.

Usando recursos computacionais para descobrir quando é que a curva atingirá os 1.5 Exabytes temos que será aproximadamente no mês 200, ou seja como o estado atual se encontra no mês 180 ($200-180=20$), daqui a 20 meses ou em Dezembro de 2017.

22 Conclusão

Com este trabalho prático, podemos conhecer várias estatísticas relevantes em relação à monitorização de redes e serviços de forma a ganhar um conhecimento mais aprofundado e assim poder prever e solucionar mais facilmente problemas futuros que possam vir a acontecer pela natural evolução da rede ou por eventos externos inesperados.