

ass2

February 17, 2020

```
[3]: import numpy as np
import re
from collections import defaultdict
import json
import os
from sklearn.preprocessing import MultiLabelBinarizer
import pandas as pd
from scipy import sparse
from networkx import convert_matrix
import networkx as nx
import matplotlib.pyplot as plt
```

```
[ ]:
```

```
[ ]:
```

```
[2]: def extract_features_for_app(apk_address):

    basepath=os.getcwd()
    ass2_json=os.path.join(basepath, 'ass2_json')
    if not os.path.exists(ass2_json):
        os.mkdir(ass2_json)

    method_pattern='^.method'
    end_method_pattern='^.end\smethod'
    app_name_pattern='.+/(.+)apk'
    target_line_pattern='invoke-.'
    invoke_method_pattern='(invoke-(direct|static|virtual|super|interface))'
    package_pattern='}, (.+)->'
    # API_call_pattern='}, (.+)$'
    API_P1=', \[?\[?(L.+;->.+)\('
    API_P2=', \[?\[?([A-Z]->.+)\('

    #app_smali_folder_dir
    API_invoke_methods=defaultdict(list)
    API_package=defaultdict(list)
```

```

app_API=defaultdict(list)
API_codeblocks=[]
all_API=[]
for cat in os.listdir(apk_address):
    cat_dir=os.path.join(apk_address, cat)
    if os.path.isdir(cat_dir):
        ##### for computation, I only take 2
        app_smali_folder_path=[os.path.join(cat_dir, smali_class) for
→smali_class in os.listdir(cat_dir)\
                                if not smali_class.endswith('.apk') and os.
→path.isdir(os.path.join(cat_dir, \
                                smali_class))]
→
    print(app_smali_folder_path)
    for app_smali_folder_dir in app_smali_folder_path:
        app_name=re.findall(app_name_pattern,app_smali_folder_dir)[0]
        print(app_name)
        add_line=False
        for dirpath, dirname, files in os.walk(app_smali_folder_dir):
            for file in files:
                if file.endswith('.smali'):
                    with open(os.path.join(dirpath, file), 'r') as auto:
                        with open(auto.name, 'r') as f:
                            for i in f.readlines():
                                start_method=re.match(method_pattern, i)
                                end_method=re.match(end_method_pattern,
→i)
                                target_line=re.
→findall(target_line_pattern, i)
                                if start_method:
                                    add_line=True
                                    API_single_codeblock=[]
                                if end_method:
                                    add_line=False
                                    API_codeblocks.
→append(API_single_codeblock)
                                if add_line==True:
                                    #extract invoke method, API calls
→once the line is target line within the code clocks
                                    if len(target_line)!=0:
                                        #extract
→invoke_method+package+API calls
                                        #method_type is among
→'invoke-static', 'invoke-virtual',
                                        #'invoke-direct',
→'invoke-super', 'invoke-interface'

```

```

        try:
            invoke_method=re.

            package=re.

            API_call_ls=re.

            API_call=API_call_ls[0]
        except:
            print(i)
            if API_call not in _

    ↪API_invoke_methods[invoke_method]:

    ↪API_invoke_methods[invoke_method].append(API_call)

    ↪API_package[package]:

    ↪append(API_call)

    ↪block

    ↪API_single_codeblock:

    ↪append(API_call)

    ↪append(API_call)

        if API_call not in all_API:
            all_API.append(API_call)

    output=[API_package,API_invoke_methods,app_API,API_codeblocks, all_API]
    output_name=['API_package', 'API_invoke_methods', 'app_API',_
    ↪'API_codeblocks', 'all_API']
#     for i in range(len(output)):
#         data=output[i]
#         with open(ass2_json+'/' +output_name[i]+' .json', 'w+') as f:
#             json.dump(data, f)
#             print(output_name[i], "->complete")
    return output

```

```
[263]: output=extract_features_for_app(apk_address)
```

```
['/datasets/dsc180a-wi20-public/Malware/testing/benign/Microsoft Word Write Edit  
Share Docs on the Go_v16.0.12430.20120_apkpure.com/smali']  
Microsoft Word Write Edit Share Docs on the Go_v16.0.12430.20120_  
['/datasets/dsc180a-wi20-public/Malware/testing/benign/Worms Zone io Voracious  
Snake_v1.2.9_apkpure.com/smali']  
Worms Zone io Voracious Snake_v1.2.9_  
['/datasets/dsc180a-wi20-public/Malware/testing/benign/Happy  
Glass_v1.0.47_apkpure.com/smali']  
Happy Glass_v1.0.47_  
['/datasets/dsc180a-wi20-public/Malware/testing/benign/Subway  
Surfers_v1.114.0_apkpure.com/smali']  
Subway Surfers_v1.114.0_  
['/datasets/dsc180a-wi20-public/Malware/testing/benign/Instagram  
Lite_v67.0.0.0.51_apkpure.com/smali']  
Instagram Lite_v67.0.0.0.51_
```

```
[264]: apk_address='/datasets/dsc180a-wi20-public/Malware/testing/benign'
```

```
[265]: apk_address='/datasets/dsc180a-wi20-public/Malware/testing/'  
os.listdir(apk_address)
```

```
[265]: ['benign', 'malware']
```

```
[266]: ass2_json='/home/sic100/ass2_json'
```

```
[267]: output_name=['API_package', 'API_invoke_methods', 'app_API', 'API_codeblocks',  
↳ 'all_API']  
for i in range(len(output)):  
    data=output[i]  
    with open(ass2_json+'/'+output_name[i]+'.json', 'w+') as f:  
        json.dump(data, f)  
        print(output_name[i], "->complete")
```

```
API_package ->complete  
API_invoke_methods ->complete  
app_API ->complete  
API_codeblocks ->complete  
all_API ->complete
```

```
[6]: path='./ass2_json/API_package'+'.json'  
with open(path, 'r') as inputfile:  
    API_package=json.load(inputfile)  
path='./ass2_json/API_invoke_methods'+'.json'  
with open(path, 'r') as inputfile:  
    API_invoke_methods=json.load(inputfile)
```

```

path='./ass2_json/app_API'+'.json'
with open(path, 'r') as inputfile:
    app_API=json.load(inputfile)
path='./ass2_json/API_codeblocks'+'.json'
with open(path, 'r') as inputfile:
    API_codeblocks=json.load(inputfile)
path='./ass2_json/all_API'+'.json'
with open(path, 'r') as inputfile:
    all_API=json.load(inputfile)

```

```

[73]: path='./ass2_mal_json/API_package'+'.json'
with open(path, 'r') as inputfile:
    API_package_m=json.load(inputfile)
path='./ass2_mal_json/API_invoke_methods'+'.json'
with open(path, 'r') as inputfile:
    API_invoke_methods_m=json.load(inputfile)
path='./ass2_mal_json/app_API'+'.json'
with open(path, 'r') as inputfile:
    app_API_m=json.load(inputfile)
path='./ass2_mal_json/API_codeblocks'+'.json'
with open(path, 'r') as inputfile:
    API_codeblocks_m=json.load(inputfile)
path='./ass2_mal_json/all_API'+'.json'
with open(path, 'r') as inputfile:
    all_API_m=json.load(inputfile)

```

```

[3]: def get_I(API_invoke_methods, all_API):
    invoke_methods=list(API_invoke_methods.keys())
    df=pd.DataFrame(0, columns=invoke_methods, index=all_API)
    for method in invoke_methods:
        df.loc[API_invoke_methods[method], method]+=1
    return df

```

```

[4]: I=get_I(API_invoke_methods, all_API)

```

```

[5]: I.head()

```

	invoke-direct	invoke-virtual	\
Ljava/text/SimpleDateFormat;-><init>	1	0	
Ljava/lang/String;->startsWith	0	1	
Ljava/lang/StringBuilder;-><init>	1	0	
Ljava/lang/StringBuilder;->append	0	1	
Ljava/lang/StringBuilder;->toString	0	1	

	invoke-static	invoke-interface	\
Ljava/text/SimpleDateFormat;-><init>	0	0	
Ljava/lang/String;->startsWith	0	0	

Ljava/lang/StringBuilder;-><init>	0	0
Ljava/lang/StringBuilder;->append	0	0
Ljava/lang/StringBuilder;->toString	0	0
	invoke-super	
Ljava/text/SimpleDateFormat;-><init>	0	
Ljava/lang/String;->startsWith	0	
Ljava/lang/StringBuilder;-><init>	0	
Ljava/lang/StringBuilder;->append	0	
Ljava/lang/StringBuilder;->toString	0	

```
[12]: def get_A(all_API, app_API):
    app_api_df=pd.DataFrame({'app': list(app_API.keys()), 'api list':
    ↪list(app_API.values())})
    mlb = MultiLabelBinarizer()
    mlb.fit([all_API])
    one_hot=mlb.transform(app_api_df['api list'])
    A=pd.DataFrame(index=list(app_API.keys()), columns=all_API, data=one_hot)
    return A
```

```
[13]: A=get_A(all_API,app_API)
A.shape
```

```
[13]: (5, 113589)
```

```
[87]: def get_AAT(all_API, app_API):
    A=get_A(all_API, app_API)
    sparse_A=sparse.csr_matrix(A.values)
    AAT_value=sparse_A.dot(sparse_A.T).todense()
    all_app=list(app_API.keys())
    AAT=pd.DataFrame(index=all_app, columns=all_app, data=AAT_value)
    return AAT
```

```
[88]: AAT=get_AAT(all_API, app_API)
```

```
[89]: AAT
```

```
[89]: Microsoft Word Write Edit
```

```
Share Docs on the Go_v16.0.12430.20120_ \
Microsoft Word Write Edit Share Docs on the Go_...
34461
Worms Zone io Voracious Snake_v1.2.9_
5234
Happy Glass_v1.0.47_
4467
Subway Surfers_v1.114.0_
5509
```

Instagram Lite_v67.0.0.0.51_
2382

Worms Zone io Voracious

Snake_v1.2.9_ \\
Microsoft Word Write Edit Share Docs on the Go_...
5234
Worms Zone io Voracious Snake_v1.2.9_
35899
Happy Glass_v1.0.47_
11836
Subway Surfers_v1.114.0_
12430
Instagram Lite_v67.0.0.0.51_
2317

Happy Glass_v1.0.47_ \

Microsoft Word Write Edit Share Docs on the Go_... 4467
Worms Zone io Voracious Snake_v1.2.9_ 11836
Happy Glass_v1.0.47_ 35973
Subway Surfers_v1.114.0_ 10289
Instagram Lite_v67.0.0.0.51_ 1837

Subway Surfers_v1.114.0_ \

Microsoft Word Write Edit Share Docs on the Go_... 5509
Worms Zone io Voracious Snake_v1.2.9_ 12430
Happy Glass_v1.0.47_ 10289
Subway Surfers_v1.114.0_ 37173
Instagram Lite_v67.0.0.0.51_ 2014

Instagram Lite_v67.0.0.0.51_

Microsoft Word Write Edit Share Docs on the Go_... 2382
Worms Zone io Voracious Snake_v1.2.9_ 2317
Happy Glass_v1.0.47_ 1837
Subway Surfers_v1.114.0_ 2014
Instagram Lite_v67.0.0.0.51_ 8313

```
[7]: def get_B(API_codeblocks, all_API):  
    two_more=[x for x in API_codeblocks if len(x)>=2]  
    print(len(two_more))  
    mlb = MultiLabelBinarizer()  
    mlb.fit([all_API])  
    one_hot=mlb.transform(two_more)  
    codeblock_df_t=pd.DataFrame(index=range(len(two_more)), columns=all_API,  
    ↪data=one_hot)  
    print(codeblock_df_t.shape)  
    sparse_t=sparse.csr_matrix(codeblock_df_t.values)
```

```
df_data=sparse_t.T.dot(sparse_t).todense()
print(df_data.shape)
B=pd.DataFrame(index=all_API, columns=all_API, data=df_data)
return B
```

```
[8]: B=get_B(API_codeblocks, all_API)
```

```
88345
(88345, 113589)
(113589, 113589)
```

```
[9]: B.head()
```

```
[9]:                                     Ljava/text/SimpleDateFormat;-><init>  \
Ljava/text/SimpleDateFormat;-><init>                                     12
Ljava/lang/String;->startsWith                                           0
Ljava/lang/StringBuilder;-><init>                                         4
Ljava/lang/StringBuilder;->append                                         4
Ljava/lang/StringBuilder;->toString                                       4

                                     Ljava/lang/String;->startsWith  \
Ljava/text/SimpleDateFormat;-><init>                                     0
Ljava/lang/String;->startsWith                                           0
Ljava/lang/StringBuilder;-><init>                                         0
Ljava/lang/StringBuilder;->append                                         0
Ljava/lang/StringBuilder;->toString                                       0

                                     Ljava/lang/StringBuilder;-><init>  \
Ljava/text/SimpleDateFormat;-><init>                                     4
Ljava/lang/String;->startsWith                                           0
Ljava/lang/StringBuilder;-><init>                                         4
Ljava/lang/StringBuilder;->append                                         4
Ljava/lang/StringBuilder;->toString                                       4

                                     Ljava/lang/StringBuilder;->append  \
Ljava/text/SimpleDateFormat;-><init>                                     4
Ljava/lang/String;->startsWith                                           0
Ljava/lang/StringBuilder;-><init>                                         4
Ljava/lang/StringBuilder;->append                                         4
Ljava/lang/StringBuilder;->toString                                       4

                                     Ljava/lang/StringBuilder;->toString  \
Ljava/text/SimpleDateFormat;-><init>                                     4
Ljava/lang/String;->startsWith                                           0
Ljava/lang/StringBuilder;-><init>                                         4
Ljava/lang/StringBuilder;->append                                         4
Ljava/lang/StringBuilder;->toString                                       11
```



```

Lcom/net/url/URLConnection;->encode \
Ljava/text/SimpleDateFormat;-><init> 4
Ljava/lang/String;->startsWith 0
Ljava/lang/StringBuilder;-><init> 4
Ljava/lang/StringBuilder;->append 4
Ljava/lang/StringBuilder;->toString 4

Ljava/lang/String;->replace \
Ljava/text/SimpleDateFormat;-><init> 0
Ljava/lang/String;->startsWith 0
Ljava/lang/StringBuilder;-><init> 0
Ljava/lang/StringBuilder;->append 0
Ljava/lang/StringBuilder;->toString 0

Lcom/dropbox/client2/a;->a \
Ljava/text/SimpleDateFormat;-><init> 0
Ljava/lang/String;->startsWith 0
Ljava/lang/StringBuilder;-><init> 0
Ljava/lang/StringBuilder;->append 0
Ljava/lang/StringBuilder;->toString 0

Ljava/lang/IllegalArgumentException;-><init> \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Ljava/lang/Object;-><init> ... \
Ljava/text/SimpleDateFormat;-><init> 0 ...
Ljava/lang/String;->startsWith 0 ...
Ljava/lang/StringBuilder;-><init> 0 ...
Ljava/lang/StringBuilder;->append 0 ...
Ljava/lang/StringBuilder;->toString 0 ...

Landroidx/slidingpanelayout/widget/SlidingPanelLayout;->a \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>

```

```

0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->b \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/a;-><init> \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->d \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/b;-><init> \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0

```

```

Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

                                Landroidx/slidingpanelayout/widget/a;->run
\
Ljava/text/SimpleDateFormat;-><init>                                0
Ljava/lang/String;->startsWith                                      0
Ljava/lang/StringBuilder;-><init>                                    0
Ljava/lang/StringBuilder;->append                                  0
Ljava/lang/StringBuilder;->toString                                0

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->c  \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/SlidingPaneLayout$SavedState;->a  \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->e  \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append

```

```

0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/d;-><init>
Ljava/text/SimpleDateFormat;-><init>
4
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
4
Ljava/lang/StringBuilder;->append
4
Ljava/lang/StringBuilder;->toString
4

```

[5 rows x 113589 columns]

```

[39]: def get_P(API_package, all_API):
        mlb = MultiLabelBinarizer()
        mlb.fit([all_API])
        one_hot=mlb.transform(list(API_package.values()))
        df_t=pd.DataFrame(columns=all_API, index=list(API_package.keys()),
        ↪data=one_hot)
        sparse_t=sparse.csr_matrix(df_t.values)
        df_data=sparse_t.T.dot(sparse_t).todense()
        P=pd.DataFrame(index=all_API, columns=all_API, data=df_data)
        return P

```

```

[15]: P=get_P(API_package, all_API)

```

```

(28039, 113589)
(113589, 113589)

```

```

[17]: P.head()

```

```

[17]:
        Ljava/text/SimpleDateFormat;-><init> \
Ljava/text/SimpleDateFormat;-><init>                2
Ljava/lang/String;->startsWith                      0
Ljava/lang/StringBuilder;-><init>                    0
Ljava/lang/StringBuilder;->append                   0
Ljava/lang/StringBuilder;->toString                 0

        Ljava/lang/String;->startsWith \
Ljava/text/SimpleDateFormat;-><init>                0
Ljava/lang/String;->startsWith                      1
Ljava/lang/StringBuilder;-><init>                    0

```

Ljava/lang/StringBuilder;->append	0	
Ljava/lang/StringBuilder;->toString	0	
		Ljava/lang/StringBuilder;-<init> \
Ljava/text/SimpleDateFormat;-<init>	0	
Ljava/lang/String;->startsWith	0	
Ljava/lang/StringBuilder;-<init>	1	
Ljava/lang/StringBuilder;->append	0	
Ljava/lang/StringBuilder;->toString	0	
		Ljava/lang/StringBuilder;->append \
Ljava/text/SimpleDateFormat;-<init>	0	
Ljava/lang/String;->startsWith	0	
Ljava/lang/StringBuilder;-<init>	0	
Ljava/lang/StringBuilder;->append	1	
Ljava/lang/StringBuilder;->toString	0	
		Ljava/lang/StringBuilder;->toString \
Ljava/text/SimpleDateFormat;-<init>	0	
Ljava/lang/String;->startsWith	0	
Ljava/lang/StringBuilder;-<init>	0	
Ljava/lang/StringBuilder;->append	0	
Ljava/lang/StringBuilder;->toString	1	
		Ljava/net/URLDecoder;->encode \
Ljava/text/SimpleDateFormat;-<init>	0	
Ljava/lang/String;->startsWith	0	
Ljava/lang/StringBuilder;-<init>	0	
Ljava/lang/StringBuilder;->append	0	
Ljava/lang/StringBuilder;->toString	0	
		Ljava/lang/String;->replace \
Ljava/text/SimpleDateFormat;-<init>	0	
Ljava/lang/String;->startsWith	0	
Ljava/lang/StringBuilder;-<init>	0	
Ljava/lang/StringBuilder;->append	0	
Ljava/lang/StringBuilder;->toString	0	
		Lcom/dropbox/client2/a;->a \
Ljava/text/SimpleDateFormat;-<init>	0	
Ljava/lang/String;->startsWith	0	
Ljava/lang/StringBuilder;-<init>	0	
Ljava/lang/StringBuilder;->append	0	
Ljava/lang/StringBuilder;->toString	0	
		Ljava/lang/IllegalArgumentException;-<init> \
Ljava/text/SimpleDateFormat;-<init>		

```

0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

                                Ljava/lang/Object;-><init>    ... \
Ljava/text/SimpleDateFormat;-><init>                        0 ...
Ljava/lang/String;->startsWith                             0 ...
Ljava/lang/StringBuilder;-><init>                           0 ...
Ljava/lang/StringBuilder;->append                          0 ...
Ljava/lang/StringBuilder;->toString                        0 ...

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->a \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->b \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/a;-><init> \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>

```

```

0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->d \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

Landroidx/slidingpanelayout/widget/b;-><init> \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0
Ljava/lang/StringBuilder;->toString
0

\
Landroidx/slidingpanelayout/widget/a;->run
\
Ljava/text/SimpleDateFormat;-><init> 0
Ljava/lang/String;->startsWith 0
Ljava/lang/StringBuilder;-><init> 0
Ljava/lang/StringBuilder;->append 0
Ljava/lang/StringBuilder;->toString 0

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->c \
Ljava/text/SimpleDateFormat;-><init>
0
Ljava/lang/String;->startsWith
0
Ljava/lang/StringBuilder;-><init>
0
Ljava/lang/StringBuilder;->append
0

```

```
Ljava/lang/StringBuilder;->toString
0
```

```
Landroidx/slidingpanelayout/widget/SlidingPaneLayout$SavedState;->a \
Ljava/text/SimpleDateFormat;-><init>
```

```
0
Ljava/lang/String;->startsWith
```

```
0
Ljava/lang/StringBuilder;-><init>
```

```
0
Ljava/lang/StringBuilder;->append
```

```
0
Ljava/lang/StringBuilder;->toString
```

```
0
```

```
Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->e \
```

```
Ljava/text/SimpleDateFormat;-><init>
```

```
0
Ljava/lang/String;->startsWith
```

```
0
Ljava/lang/StringBuilder;-><init>
```

```
0
Ljava/lang/StringBuilder;->append
```

```
0
Ljava/lang/StringBuilder;->toString
```

```
0
```

```
Landroidx/slidingpanelayout/widget/d;-><init>
```

```
Ljava/text/SimpleDateFormat;-><init>
```

```
0
Ljava/lang/String;->startsWith
```

```
0
Ljava/lang/StringBuilder;-><init>
```

```
0
Ljava/lang/StringBuilder;->append
```

```
0
Ljava/lang/StringBuilder;->toString
```

```
0
```

```
[5 rows x 113589 columns]
```

```
[ ]:
```

```
[93]: info=AAT.stack()[AAT.stack()>0]
df=info.to_frame('value').reset_index()
df
```


[93]:

```
level_0 \
0  Microsoft Word Write Edit Share Docs on the Go...
1  Microsoft Word Write Edit Share Docs on the Go...
2  Microsoft Word Write Edit Share Docs on the Go...
3  Microsoft Word Write Edit Share Docs on the Go...
4  Microsoft Word Write Edit Share Docs on the Go...
5      Worms Zone io Voracious Snake_v1.2.9_
6      Worms Zone io Voracious Snake_v1.2.9_
7      Worms Zone io Voracious Snake_v1.2.9_
8      Worms Zone io Voracious Snake_v1.2.9_
9      Worms Zone io Voracious Snake_v1.2.9_
10     Happy Glass_v1.0.47_
11     Happy Glass_v1.0.47_
12     Happy Glass_v1.0.47_
13     Happy Glass_v1.0.47_
14     Happy Glass_v1.0.47_
15     Subway Surfers_v1.114.0_
16     Subway Surfers_v1.114.0_
17     Subway Surfers_v1.114.0_
18     Subway Surfers_v1.114.0_
19     Subway Surfers_v1.114.0_
20     Instagram Lite_v67.0.0.0.51_
21     Instagram Lite_v67.0.0.0.51_
22     Instagram Lite_v67.0.0.0.51_
23     Instagram Lite_v67.0.0.0.51_
24     Instagram Lite_v67.0.0.0.51_

level_1  value
0  Microsoft Word Write Edit Share Docs on the Go... 34461
1      Worms Zone io Voracious Snake_v1.2.9_ 5234
2          Happy Glass_v1.0.47_ 4467
3      Subway Surfers_v1.114.0_ 5509
4      Instagram Lite_v67.0.0.0.51_ 2382
5  Microsoft Word Write Edit Share Docs on the Go... 5234
6      Worms Zone io Voracious Snake_v1.2.9_ 35899
7          Happy Glass_v1.0.47_ 11836
8      Subway Surfers_v1.114.0_ 12430
9      Instagram Lite_v67.0.0.0.51_ 2317
10 Microsoft Word Write Edit Share Docs on the Go... 4467
11     Worms Zone io Voracious Snake_v1.2.9_ 11836
12         Happy Glass_v1.0.47_ 35973
13         Subway Surfers_v1.114.0_ 10289
14         Instagram Lite_v67.0.0.0.51_ 1837
15 Microsoft Word Write Edit Share Docs on the Go... 5509
16     Worms Zone io Voracious Snake_v1.2.9_ 12430
17         Happy Glass_v1.0.47_ 10289
18         Subway Surfers_v1.114.0_ 37173
```

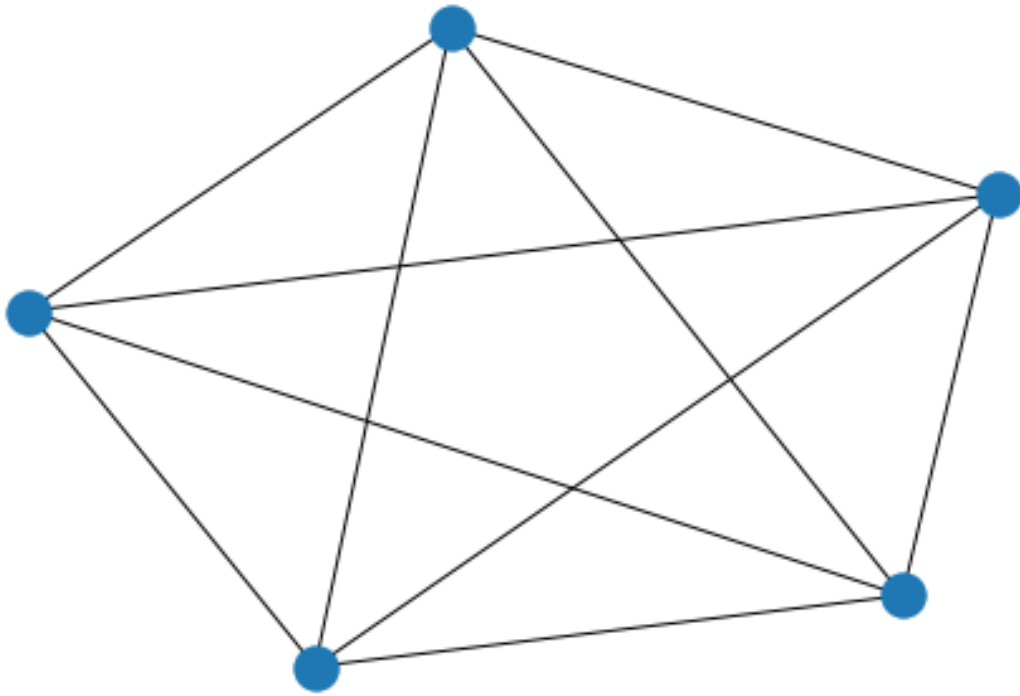
19	Instagram Lite_v67.0.0.0.51_	2014
20	Microsoft Word Write Edit Share Docs on the Go...	2382
21	Worms Zone io Voracious Snake_v1.2.9_	2317
22	Happy Glass_v1.0.47_	1837
23	Subway Surfers_v1.114.0_	2014
24	Instagram Lite_v67.0.0.0.51_	8313

```
[94]: df.level_0.tolist()
```

```
[94]: ['Microsoft Word Write Edit Share Docs on the Go_v16.0.12430.20120_',
'Microsoft Word Write Edit Share Docs on the Go_v16.0.12430.20120_',
'Microsoft Word Write Edit Share Docs on the Go_v16.0.12430.20120_',
'Microsoft Word Write Edit Share Docs on the Go_v16.0.12430.20120_',
'Microsoft Word Write Edit Share Docs on the Go_v16.0.12430.20120_',
'Worms Zone io Voracious Snake_v1.2.9_',
'Worms Zone io Voracious Snake_v1.2.9_',
'Worms Zone io Voracious Snake_v1.2.9_',
'Worms Zone io Voracious Snake_v1.2.9_',
'Worms Zone io Voracious Snake_v1.2.9_',
'Happy Glass_v1.0.47_',
'Happy Glass_v1.0.47_',
'Happy Glass_v1.0.47_',
'Happy Glass_v1.0.47_',
'Happy Glass_v1.0.47_',
'Subway Surfers_v1.114.0_',
'Subway Surfers_v1.114.0_',
'Subway Surfers_v1.114.0_',
'Subway Surfers_v1.114.0_',
'Subway Surfers_v1.114.0_',
'Instagram Lite_v67.0.0.0.51_',
'Instagram Lite_v67.0.0.0.51_',
'Instagram Lite_v67.0.0.0.51_',
'Instagram Lite_v67.0.0.0.51_',
'Instagram Lite_v67.0.0.0.51_']
```

```
[95]: G=nx.from_pandas_edgelist(df, 'level_0','level_1', ['value'])
```

```
[99]: Gr=nx.draw(G)
```



```
[ ]:
```

```
[68]: def draw_graph(matrix):
        s=matrix.stack()[matrix.stack()>0]
        df=s.to_frame('value').reset_index()
        G=nx.from_pandas_edgelist(df, 'level_0','level_1', ['value'])
        nx.draw(G)
        return G
```

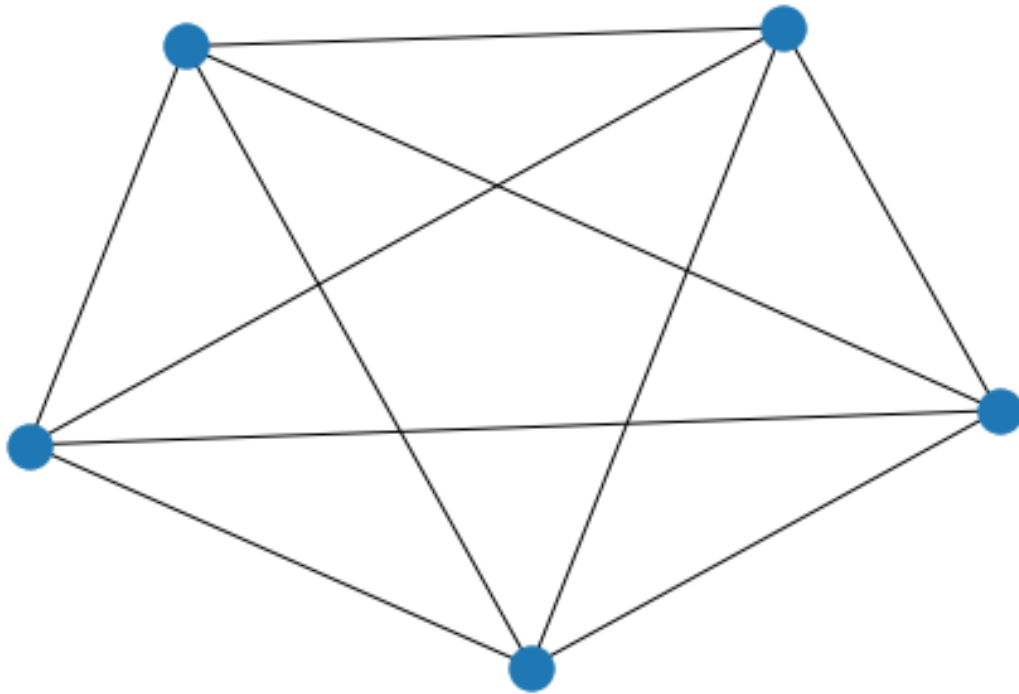
```
[69]: def graph_stats(G):
        density = nx.density(G)
        corr = nx.degree_pearson_correlation_coefficient(G)
        avg_neighbor_degree = nx.average_neighbor_degree(G) #dict
        k_nearest_neighbors = nx.k_nearest_neighbors(G) #dict
        degree centrality = nx.degree_centrality(G) #dict
        info = nx.info(G)
        return {'density':density, 'corr':corr, 'avg_neighbor_degree':
        ↪avg_neighbor_degree,
                'k_nearest_neighbors':k_nearest_neighbors, 'degree_centrality':
        ↪degree_centrality, 'info':info}
```

```
[70]: G_ABAT=draw_graph(ABAT)
```

```

/opt/conda/lib/python3.7/site-packages/networkx/drawing/nx_pylab.py:579:
MatplotlibDeprecationWarning:
The iterable function was deprecated in Matplotlib 3.1 and will be removed in
3.3. Use np.iterable instead.
    if not cb.iterable(width):

```



```

[72]: stars_ABAT=graph_stats(G_ABAT)
      stars_ABAT

```

```

[72]: {'density': 1.5,
      'corr': nan,
      'avg_neighbor_degree': {'Microsoft Word Write Edit Share Docs on the
Go_v16.0.12430.20120_': 5.0,
      'Worms Zone io Voracious Snake_v1.2.9_': 5.0,
      'Happy Glass_v1.0.47_': 5.0,
      'Subway Surfers_v1.114.0_': 5.0,
      'Instagram Lite_v67.0.0.0.51_': 5.0},
      'k_nearest_neighbors': {6: 5.0},
      'degree_centrality': {'Microsoft Word Write Edit Share Docs on the
Go_v16.0.12430.20120_': 1.5,
      'Worms Zone io Voracious Snake_v1.2.9_': 1.5,
      'Happy Glass_v1.0.47_': 1.5,
      'Subway Surfers_v1.114.0_': 1.5,

```

```

    'Instagram Lite_v67.0.0.0.51_': 1.5},
    'info': 'Name: \nType: Graph\nNumber of nodes: 5\nNumber of edges: 15\nAverage
degree: 6.0000'}

```

```

[34]: def get_ABAT(API_codeblocks, all_API, app_API):
        all_app=list(app_API.keys())
        A_values=get_A(all_API, app_API).values
        B_values=get_B(API_codeblocks, all_API).values
        sparse_A=sparse.csr_matrix(A_values)
        sparse_B=sparse.csr_matrix(B_values)
        sparse_B[sparse_B>0]=1
        result=sparse_A.dot(sparse_B).dot(sparse_A.T).todense()
        ABAT=pd.DataFrame(index=all_app, columns=all_app, data=result)
        return ABAT

```

```

[35]: ABAT=get_ABAT(API_codeblocks, all_API, app_API)

```

```

88345
(88345, 113589)
(113589, 113589)

```

```

[36]: ABAT

```

```

[36]: Microsoft Word Write Edit

```

```

Share Docs on the Go_v16.0.12430.20120_ \
Microsoft Word Write Edit Share Docs on the Go_...
867398
Worms Zone io Voracious Snake_v1.2.9_
520734
Happy Glass_v1.0.47_
538883
Subway Surfers_v1.114.0_
547471
Instagram Lite_v67.0.0.0.51_
331071

```

Worms Zone io Voracious

```

Snake_v1.2.9_ \
Microsoft Word Write Edit Share Docs on the Go_...
520734
Worms Zone io Voracious Snake_v1.2.9_
962310
Happy Glass_v1.0.47_
665244
Subway Surfers_v1.114.0_
692424
Instagram Lite_v67.0.0.0.51_

```

336372

	Happy Glass_v1.0.47_ \
Microsoft Word Write Edit Share Docs on the Go_...	538883
Worms Zone io Voracious Snake_v1.2.9_	665244
Happy Glass_v1.0.47_	933212
Subway Surfers_v1.114.0_	657897
Instagram Lite_v67.0.0.0.51_	338566
	Subway Surfers_v1.114.0_ \
Microsoft Word Write Edit Share Docs on the Go_...	547471
Worms Zone io Voracious Snake_v1.2.9_	692424
Happy Glass_v1.0.47_	657897
Subway Surfers_v1.114.0_	1027990
Instagram Lite_v67.0.0.0.51_	347802
	Instagram Lite_v67.0.0.0.51_
Microsoft Word Write Edit Share Docs on the Go_...	331071
Worms Zone io Voracious Snake_v1.2.9_	336372
Happy Glass_v1.0.47_	338566
Subway Surfers_v1.114.0_	347802
Instagram Lite_v67.0.0.0.51_	266120

```
[14]: A_values=A.values
```

```
[15]: sparse_A=sparse.csr_matrix(A_values)
```

```
[29]: result=sparse_A.dot(sparse_B).dot(sparse_A.T).todense()
```

```
[30]: result
```

```
[30]: matrix([[ 867398,  520734,  538883,  547471,  331071],
 [ 520734,  962310,  665244,  692424,  336372],
 [ 538883,  665244,  933212,  657897,  338566],
 [ 547471,  692424,  657897, 1027990,  347802],
 [ 331071,  336372,  338566,  347802, 266120]], dtype=int64)
```

```
[37]: def get_APAT(API_package, all_API, app_API):
    all_app=list(app_API.keys())
    A_values=get_A(all_API, app_API).values
    P_values=get_P(API_package, all_API).values
    sparse_A=sparse.csr_matrix(A_values)
    sparse_P=sparse.csr_matrix(P_values)
    result=sparse_A.dot(sparse_P).dot(sparse_A.T).todense()
    APAT=pd.DataFrame(index=all_app, columns=all_app, data=result)
    return APAT
```

```
[54]: p=get_P(API_package, all_API)
```

```
[ ]:
```

```
[40]: APAT=get_APAT(API_package, all_API, app_API)
```

```
[42]: APAT.values
```

```
[42]: array([[1443759, 180619, 259381, 265457, 77932],
          [ 180619, 733339, 309303, 260381, 135706],
          [ 259381, 309303, 979201, 361338, 71428],
          [ 265457, 260381, 361338, 698503, 75614],
          [ 77932, 135706, 71428, 75614, 118495]], dtype=int64)
```

```
[43]: sparse_A=sparse.csr_matrix(A_values)
```

```
[45]: A.describe()
```

```
[45]:      Ljava/text/SimpleDateFormat;-><init>  Ljava/lang/String;->startsWith  \
count                                     5.0          5.000000
mean                                     1.0          0.200000
std                                      0.0          0.447214
min                                      1.0          0.000000
25%                                     1.0          0.000000
50%                                     1.0          0.000000
75%                                     1.0          0.000000
max                                      1.0          1.000000
```

```
      Ljava/lang/StringBuilder;-><init>  Ljava/lang/StringBuilder;->append  \
count                                     5.000000          5.000000
mean                                     0.600000          0.600000
std                                     0.547723          0.547723
min                                     0.000000          0.000000
25%                                     0.000000          0.000000
50%                                     1.000000          1.000000
75%                                     1.000000          1.000000
max                                     1.000000          1.000000
```

```
      Ljava/lang/StringBuilder;->toString  Ljava/net/URLDecoder;->encode  \
count                                     5.0          5.0
mean                                     1.0          1.0
std                                      0.0          0.0
min                                      1.0          1.0
25%                                     1.0          1.0
```

50%	1.0	1.0
75%	1.0	1.0
max	1.0	1.0

	Ljava/lang/String;->replace	Lcom/dropbox/client2/a;->a	\
count	5.000000	5.000000	
mean	0.400000	0.400000	
std	0.547723	0.547723	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	1.000000	1.000000	
max	1.000000	1.000000	

	Ljava/lang/IllegalArgumentException;-><init>	\
count	5.000000	
mean	0.400000	
std	0.547723	
min	0.000000	
25%	0.000000	
50%	0.000000	
75%	1.000000	
max	1.000000	

	Ljava/lang/Object;-><init>	...	\
count	5.000000	...	
mean	0.400000	...	
std	0.547723	...	
min	0.000000	...	
25%	0.000000	...	
50%	0.000000	...	
75%	1.000000	...	
max	1.000000	...	

	Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->a	\
count	5.000000	
mean	0.800000	
std	0.447214	
min	0.000000	
25%	1.000000	
50%	1.000000	
75%	1.000000	
max	1.000000	

	Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->b	\
count	5.000000	
mean	0.600000	

std	0.547723
min	0.000000
25%	0.000000
50%	1.000000
75%	1.000000
max	1.000000

```

Landroidx/slidingpanelayout/widget/a;-><init> \
count      5.000000
mean       0.600000
std        0.547723
min        0.000000
25%        0.000000
50%        1.000000
75%        1.000000
max        1.000000

```

```

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->d \
count      5.000000
mean       0.600000
std        0.547723
min        0.000000
25%        0.000000
50%        1.000000
75%        1.000000
max        1.000000

```

```

Landroidx/slidingpanelayout/widget/b;-><init> \
count      5.000000
mean       0.600000
std        0.547723
min        0.000000
25%        0.000000
50%        1.000000
75%        1.000000
max        1.000000

```

```

Landroidx/slidingpanelayout/widget/a;->run \
count      5.000000
mean       0.800000
std        0.447214
min        0.000000
25%        1.000000
50%        1.000000
75%        1.000000
max        1.000000

```

```

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->c \
count                    5.000000
mean                    0.800000
std                     0.447214
min                     0.000000
25%                     1.000000
50%                     1.000000
75%                     1.000000
max                     1.000000

```

```

Landroidx/slidingpanelayout/widget/SlidingPaneLayout$SavedState;->a \
count                    5.000000
mean                    0.800000
std                     0.447214
min                     0.000000
25%                     1.000000
50%                     1.000000
75%                     1.000000
max                     1.000000

```

```

Landroidx/slidingpanelayout/widget/SlidingPaneLayout;->e \
count                    5.000000
mean                    0.200000
std                     0.447214
min                     0.000000
25%                     0.000000
50%                     0.000000
75%                     0.000000
max                     1.000000

```

```

Landroidx/slidingpanelayout/widget/d;-><init>
count                    5.000000
mean                    0.600000
std                     0.547723
min                     0.000000
25%                     0.000000
50%                     1.000000
75%                     1.000000
max                     1.000000

```

[8 rows x 113589 columns]

[]:

```

[62]: def get_APBPTAT(all_API, app_API, API_package, API_codeblocks):
        all_app=list(app_API.keys())
        A_values=get_A(all_API, app_API).values

```

```

P_values=get_P(API_package, all_API).values
B_values=get_B(API_codeblocks, all_API).values
sparse_A=sparse.csr_matrix(A_values)
sparse_P=sparse.csr_matrix(P_values)
sparse_P[sparse_P>0]=1
sparse_B=sparse.csr_matrix(B_values)
sparse_B[sparse_B>0]=1

result=sparse_A.dot(sparse_P).dot(sparse_B).dot(sparse_P.T).dot(sparse_A.T).
→todense()
APBPTAT=pd.DataFrame(index=all_app, columns=all_app, data=result)
return APBPTAT

```

```
[63]: APBPTAT=get_APBPTAT(all_API, app_API, API_package, API_codeblocks)
```

```

88345
(88345, 113589)
(113589, 113589)

```

```
[64]: APBPTAT.head()
```

```
[64]: Microsoft Word Write Edit
```

```

Share Docs on the Go_v16.0.12430.20120_ \
Microsoft Word Write Edit Share Docs on the Go_...
4582944741
Worms Zone io Voracious Snake_v1.2.9_
629481203
Happy Glass_v1.0.47_
787336293
Subway Surfers_v1.114.0_
775424940
Instagram Lite_v67.0.0.0.51_
283880639

```

Worms Zone io Voracious

```

Snake_v1.2.9_ \
Microsoft Word Write Edit Share Docs on the Go_...
629481203
Worms Zone io Voracious Snake_v1.2.9_
1137145584
Happy Glass_v1.0.47_
735543640
Subway Surfers_v1.114.0_
679390955
Instagram Lite_v67.0.0.0.51_
440013495

```

Microsoft Word Write Edit Share Docs on the Go_...	Happy Glass_v1.0.47_ \
Worms Zone io Voracious Snake_v1.2.9_	787336293
Happy Glass_v1.0.47_	735543640
Subway Surfers_v1.114.0_	1298003783
Instagram Lite_v67.0.0.0.51_	896253553
	302602635
	Subway Surfers_v1.114.0_ \
Microsoft Word Write Edit Share Docs on the Go_...	775424940
Worms Zone io Voracious Snake_v1.2.9_	679390955
Happy Glass_v1.0.47_	896253553
Subway Surfers_v1.114.0_	1026089513
Instagram Lite_v67.0.0.0.51_	294084496
	Instagram Lite_v67.0.0.0.51_
Microsoft Word Write Edit Share Docs on the Go_...	283880639
Worms Zone io Voracious Snake_v1.2.9_	440013495
Happy Glass_v1.0.47_	302602635
Subway Surfers_v1.114.0_	294084496
Instagram Lite_v67.0.0.0.51_	218218038

[]:

Statistically describe the set of applications you have collected. This will require defining your own relevant statistics (e.g. number of classes/app, most common library used, graph-degree related statistics) to investigate and understand the population of apps that you've downloaded. In particular, you should look at how these statistics differ across two relevant classes of apps (this may be one-app-type vs another-app-type; malware vs benign).

```
[85]: package_info=pd.DataFrame(index=API_package.keys(), data=[len(API_package[x])
    ↪for x in API_package], columns=['count'])
```

```
[92]: most_common_package=package_info.sort_values(by=['count'],ascending=False).
    ↪iloc[:10]
most_common_package
```

```
[92]:
```

	count
Lcom/microsoft/cognitiveservices/speech/interna...	893
Lcom/ironsource/sdk/controller/WebController;	314
Landroid/view/View;	300
Landroidx/recyclerview/widget/RecyclerView;	258
Lcom/verizon/ads/vastcontroller/VASTVideoView;	232
Landroid/support/v7/widget/RecyclerView;	219
Lcom/inmobi/media/j;	210
Lcom/microsoft/office/airspace/AirspaceScrollLa...	190
Landroidx/fragment/app/Fragment;	189
Lcom/google/android/gms/internal/ads/zzbhm;	156

[]:

[]:

[]: