graph_bipartite

October 2, 2018

In [38]:
```
import networkx as nx
from networkx import *
import matplotlib.pyplot as plt
from networkx.algorithms import bipartite
from nltk import sent_tokenize,word_tokenize
import numpy as np

import sys
sys.path.append('../..')
sys.path.append('../../utils/')
from utils import *
import operator


import fastText
model = fastText.load_model('../../../Divers_Data_Maitrise/wiki.simple/wiki.simple.bin'
```

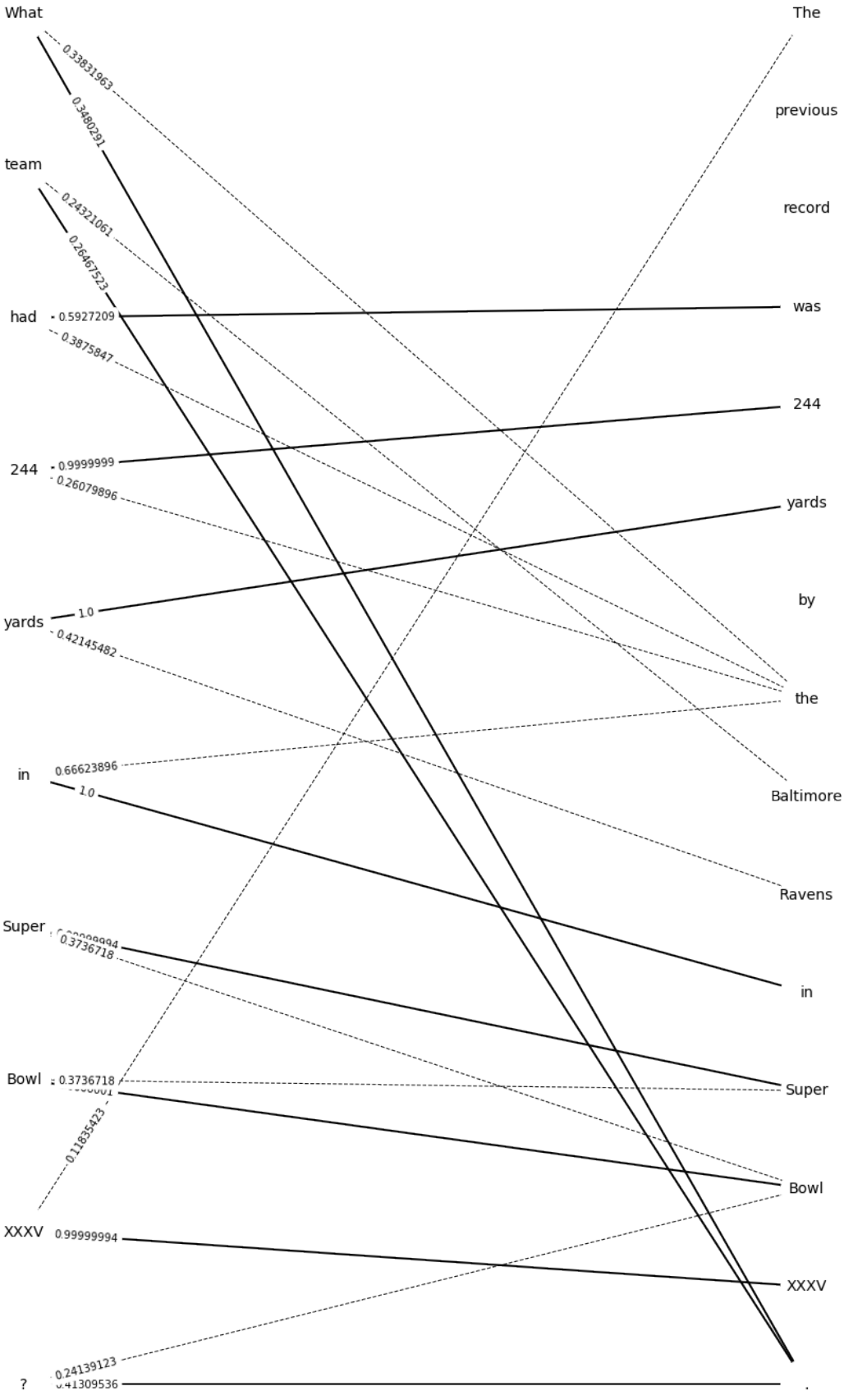
In [139]:
```
def Alignement_graph_bipartite(question,sequence, nb_alignement_a_afficher = 1 ):
    '''
    Fonction qui crée un graphe bipartit (question, sentence) et qui affecte la simila
    '''

    nb_alignement_a_afficher = max(1,nb_alignement_a_afficher)
    G=nx.Graph()

    list_words_question = word_tokenize(question)
    list_words_sequence = word_tokenize(sequence)


    nb_words_question = len(list_words_question)
    nb_words_sequence = len(list_words_sequence)
    height_colum = max(nb_words_question,nb_words_sequence)
    y_linspace_question = np.linspace(0,height_colum,nb_words_question)
    y_linspace_sentence = np.linspace(0,height_colum,nb_words_sequence)

    fixed_positions={}
    labels={}
```

```
    list_major_edges = []
    list_minor_edges = []
    edge_labels = {}
    for i in range(0,nb_words_sequence): # mots de la phrase
        num_node = nb_words_question + i
        G.add_node(num_node)
        labels[num_node]=list_words_sequence[i]

        fixed_positions[num_node] = (2,y_linspace_sentence[nb_words_sequence-i-1])

    for i in range(0,nb_words_question): # mots de la question
        G.add_node(i)
        labels[i]=list_words_question[i]
        fixed_positions[i] = (0,y_linspace_question[nb_words_question-i-1])

        max_sim = 0.0
        major_edge = None
        edges_list = []
        edges_sim_dict ={}
        vect_word_question = model.get_word_vector(list_words_question[i])
        for j in range(0,nb_words_sequence):
            num_node = nb_words_question + j
            vect_word_sentence = model.get_word_vector(list_words_sequence[j])
            sim=cosine_similarity(vect_word_sentence, vect_word_question)

            edges_sim_dict[len(edges_list)]=sim
            edges_list.append((i,num_node, sim))
            G.add_edge(i,num_node,sim=sim)

        edges_sorted = sorted(edges_sim_dict.items(), key=operator.itemgetter(1), reve
        list_major_edges.append(edges_list[edges_sorted[0][0]])
        edge_labels[(edges_list[edges_sorted[0][0]][1],edges_list[edges_sorted[0][0]][
        for k in range(1,len(edges_sorted)):
            list_minor_edges.append(edges_list[edges_sorted[k][0]])
            edge_labels[(edges_list[edges_sorted[k][0]][1],edges_list[edges_sorted[k][

    pos = nx.spring_layout(G,pos=fixed_positions, fixed=fixed_positions.keys())
    plt.figure(3,figsize=(15,height_colum*1.75))
    nx.draw_networkx_nodes(G,pos,node_color='w',node_size=2500)
    nx.draw_networkx_labels(G,pos,labels,font_size=14)
    nx.draw_networkx_edge_labels(G,pos,node_color='b',edge_labels=edge_labels, label_p
    nx.draw_networkx_edges(G,pos,node_color='b', edgelist =list_major_edges,  width =
    nx.draw_networkx_edges(G,pos,node_color='b', edgelist =list_minor_edges,style = 'd

    plt.axis('off')
#     plt.savefig("Graph.png", format="PNG") # pour enregistrer l'image dans un fichie
```

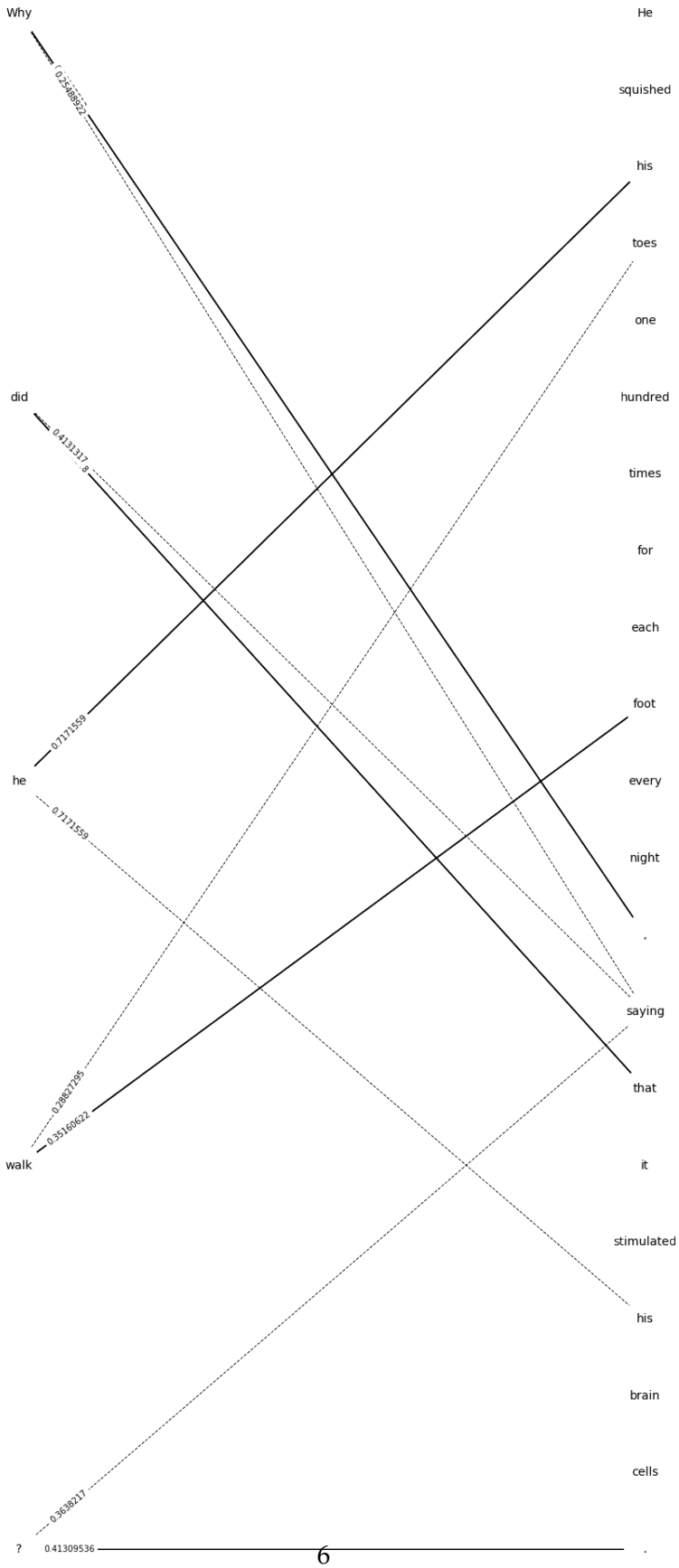
```
            plt.show()
```

```
In [141]: Alignement_graph_bipartite(" What team had 244 yards in Super Bowl XXXV? "," The previ
```

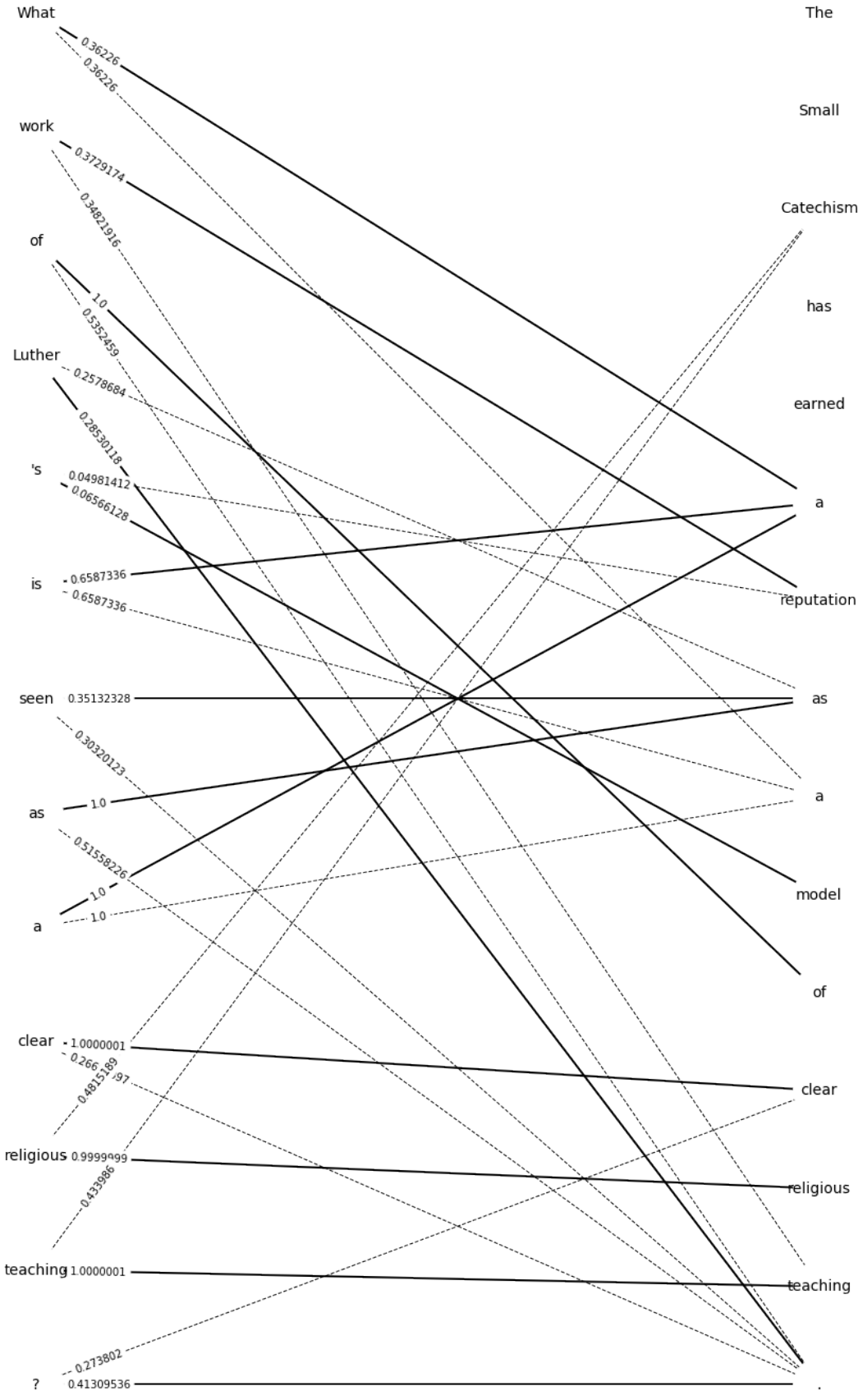
```
/media/ilan/DATA/Maitrise/env/lib/python3.6/site-packages/networkx/drawing/nx_pylab.py:611: Matp
  if cb.is_numlike(alpha):
```


What
team
had
244
yards
in
Super
Bowl
XXXV
?
The
previous
record
was
244
yards
by
the
Baltimore
Ravens
in
Super
Bowl
XXXV
.
0.33831963
0.3480291
0.24321061
0.26467523
0.5927209
0.3875847
0.9999999
0.26079896
1.0
0.42145482
0.66623896
1.0
0.3736718
0.3736718
0.11835423
0.99999994
0.24139123
0.41309536

```
In [142]: Alignement_graph_bipartite(" Why did he walk? "," He squished his toes one hundred tim

/media/ilan/DATA/Maitrise/env/lib/python3.6/site-packages/networkx/drawing/nx_pylab.py:611: Matp
  if cb.is_numlike(alpha):
```

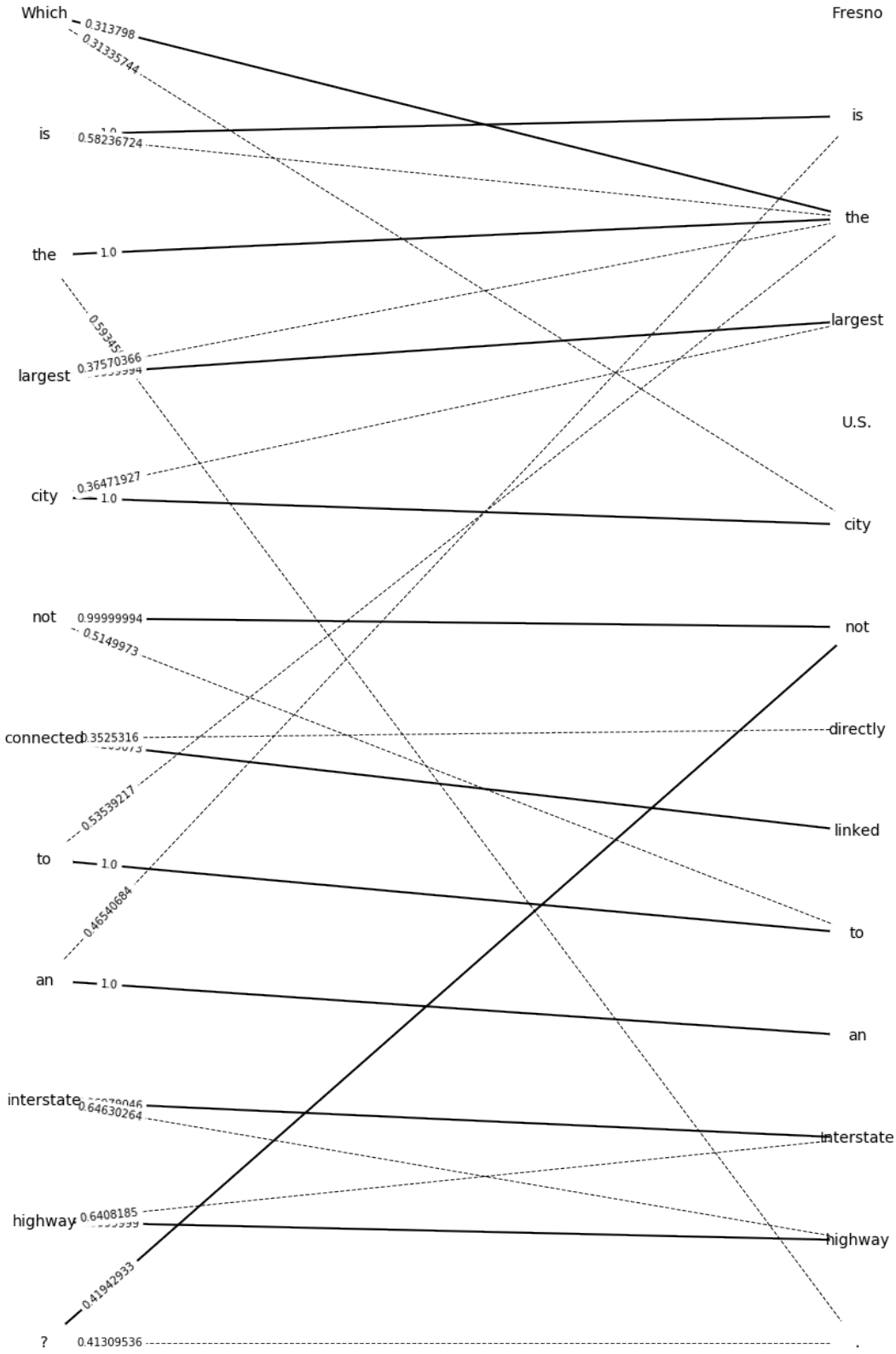

Why
did
he
walk
?
He
squished
his
toes
one
hundred
times
for
each
foot
every
night
,
saying
that
it
stimulated
his
brain
cells
.
0.41309536
0.3638217
0.28827295
0.35160622
0.7171559
0.7171559
0.4131317
0.2548892

```
In [143]: Alignement_graph_bipartite(" What work of Luther's is seen as a clear religious teachi

/media/ilan/DATA/Maitrise/env/lib/python3.6/site-packages/networkx/drawing/nx_pylab.py:611: Matp
  if cb.is_numlike(alpha):
```


What
work
of
Luther
's
is
seen
as
a
clear
religious
teaching
?
The
Small
Catechism
has
earned
a
reputation
as
a
model
of
clear
religious
teaching
.
0.36226
0.36226
0.3729174
0.34821916
1.0
0.5352459
0.2578684
0.28530118
0.04981412
0.06566128
0.6587336
0.6587336
0.35132328
0.30320123
1.0
0.51558226
1.0
1.0
1.0000001
0.26615289
0.48151797
0.9999999
0.433986
1.0000001
0.273802
0.41309536

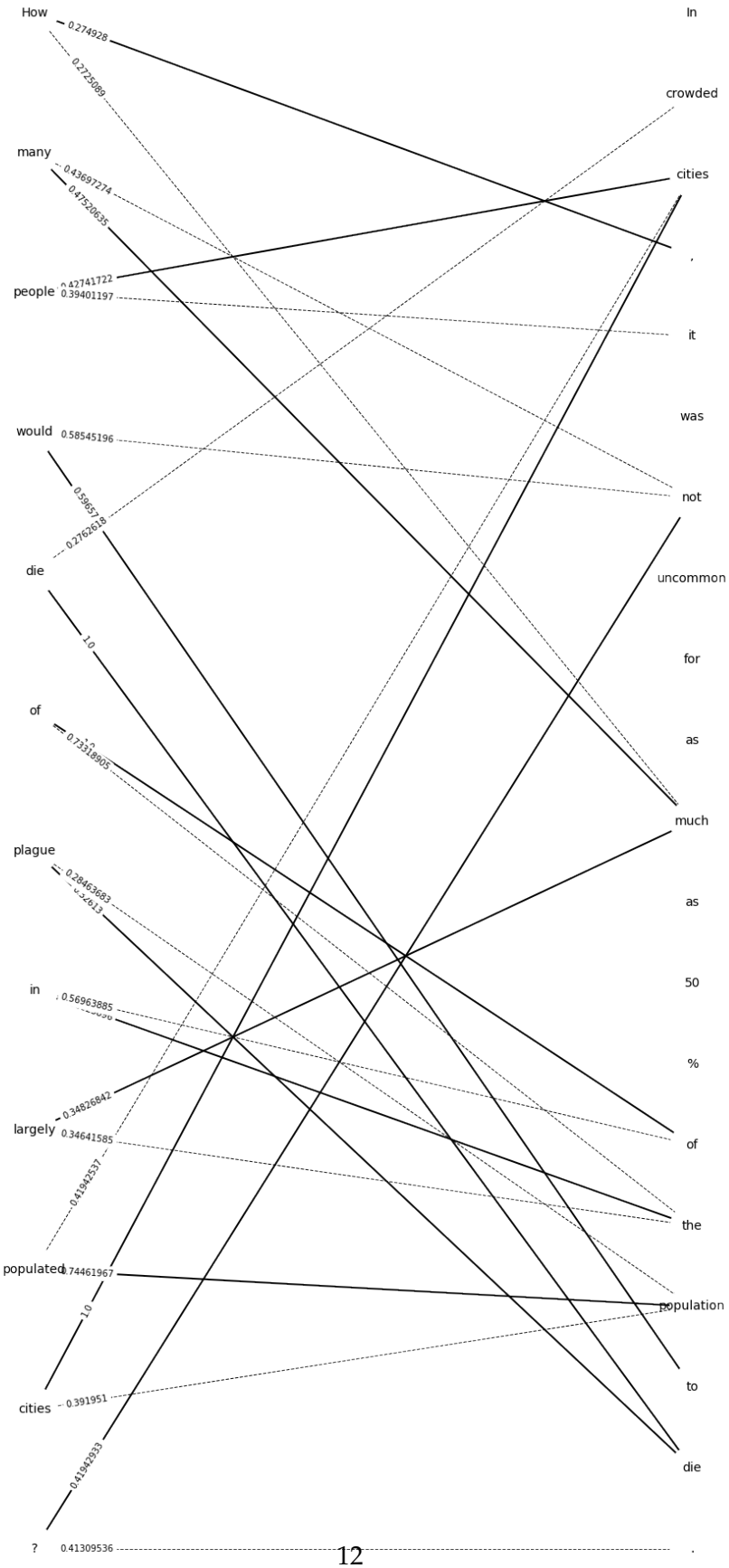
```
In [144]: Alignement_graph_bipartite(" Which is the largest city not connected to an interstate
```

```
/media/ilan/DATA/Maitrise/env/lib/python3.6/site-packages/networkx/drawing/nx_pylab.py:611: Matp
  if cb.is_numlike(alpha):
```

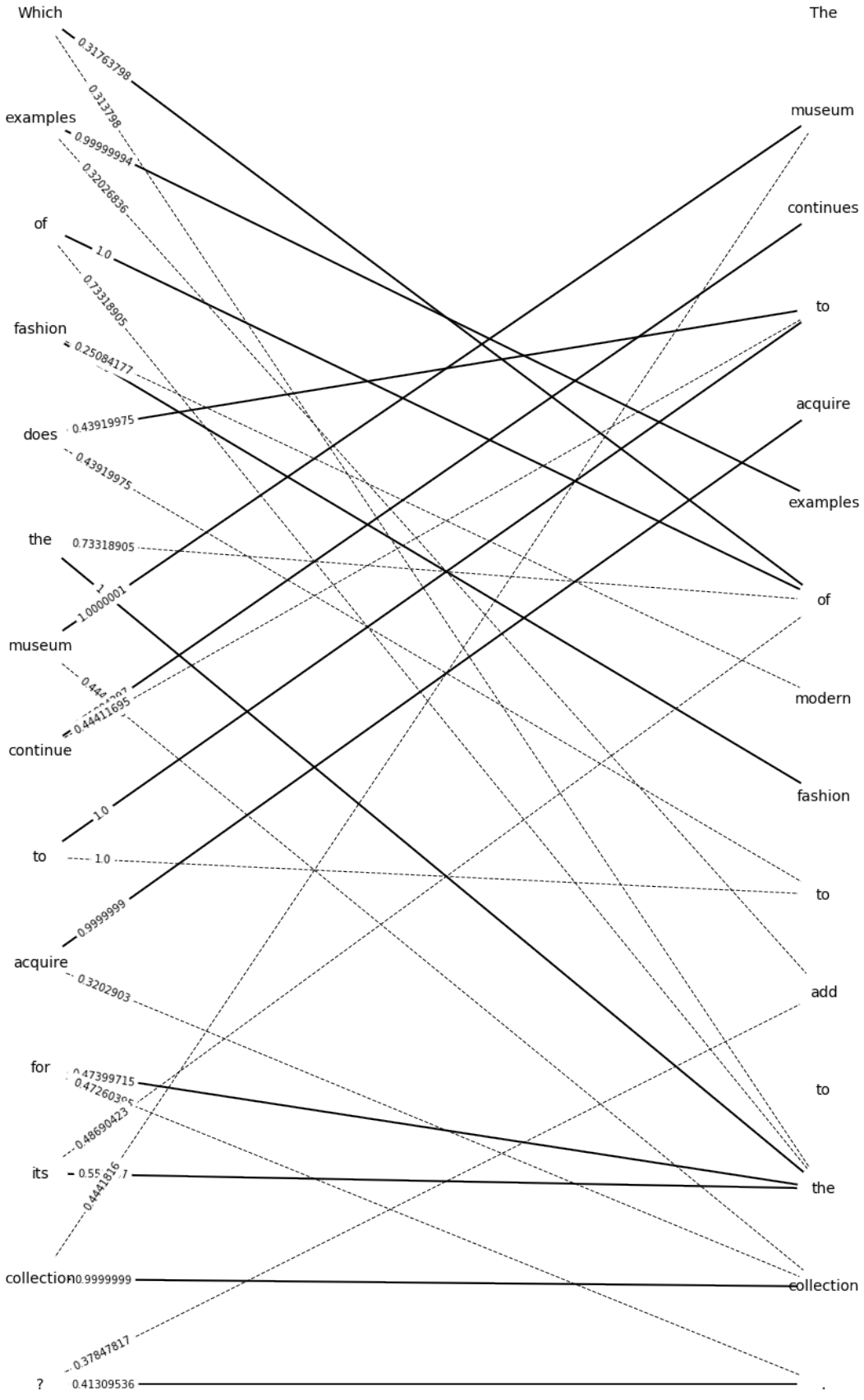

Which
is
the
largest
city
not
connected
to
an
interstate
highway
?
Fresno
is
the
largest
U.S.
city
not
directly
linked
to
an
interstate
highway
.
0.313798
0.31335744
0.58236724
1.0
0.37570366
0.36471927
1.0
0.99999994
0.5149973
0.3525316
0.53539217
1.0
0.46540684
1.0
0.64630264
0.6408185
0.41942933
0.41309536

```
In [145]: Alignement_graph_bipartite(" How many people would die of plague in largely populated
```

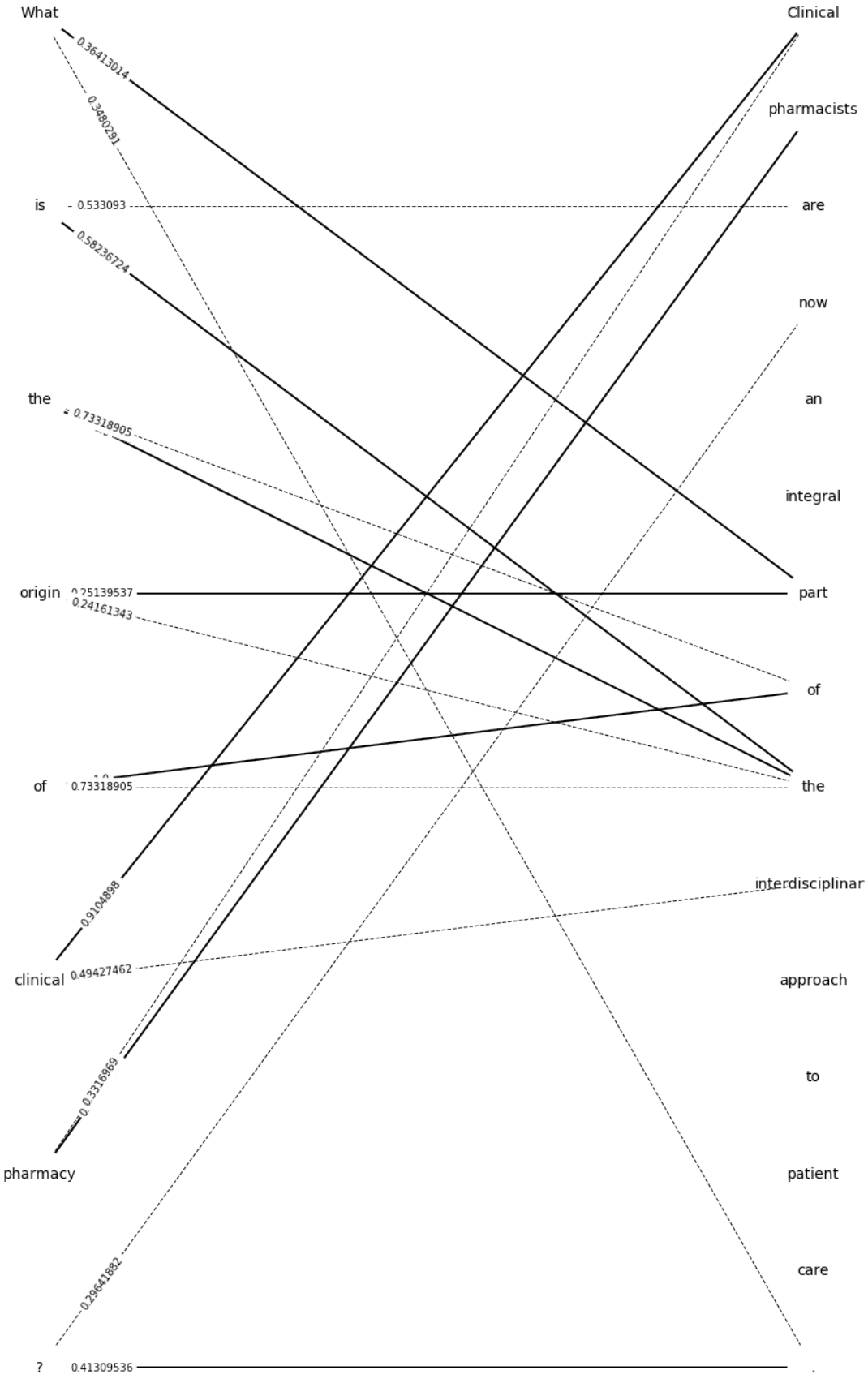
```
/media/ilan/DATA/Maitrise/env/lib/python3.6/site-packages/networkx/drawing/nx_pylab.py:611: Matp
  if cb.is_numlike(alpha):
```


How
many
people
would
die
of
plague
in
largely
populated
cities
?
In
crowded
cities
,
it
was
not
uncommon
for
as
much
as
50
%
of
the
population
to
die
.
0.274928
0.2725089
0.43697274
0.47520635
0.42741722
0.39401197
0.58545196
0.59657
0.2762618
1.0
0.73318905
0.28463683
0.56963885
0.34826842
0.34641585
0.41942537
0.74461967
1.0
0.391951
0.41942933
0.41309536

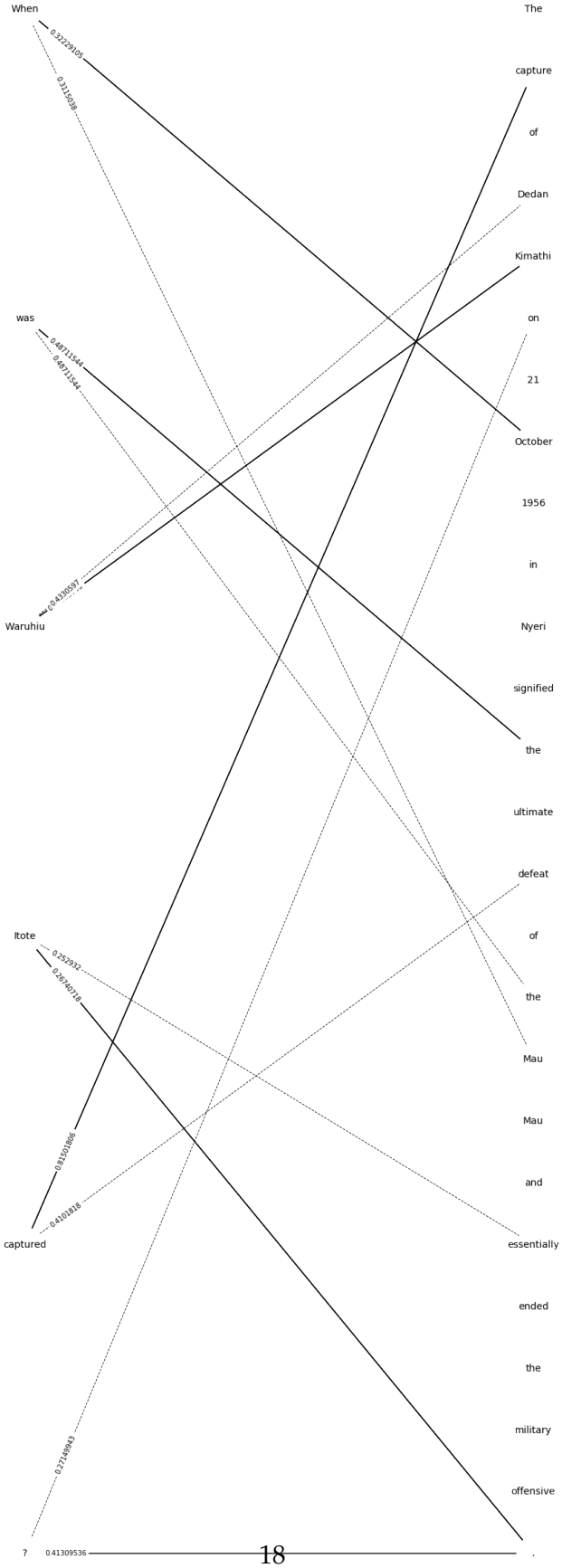
```
In [146]: Alignement_graph_bipartite(" Which examples of fashion does the museum continue to acq

/media/ilan/DATA/Maitrise/env/lib/python3.6/site-packages/networkx/drawing/nx_pylab.py:611: Matp
  if cb.is_numlike(alpha):
```


Which
examples
of
fashion
does
the
museum
continue
to
acquire
for
its
collection
?
The
museum
continues
to
acquire
examples
of
modern
fashion
to
add
to
the
collection
.
0.31763798
0.313798
0.99999994
0.32026836
1.0
0.73318905
0.25084177
0.43919975
0.43919975
0.73318905
1.0000001
0.44411695
1.0
1.0
0.9999999
0.3202903
0.47399715
0.48690423
0.4441816
0.9999999
0.37847817
0.41309536

```
In [147]: Alignement_graph_bipartite(" What is the origin of clinical pharmacy? "," Clinical pha

/media/ilan/DATA/Maitrise/env/lib/python3.6/site-packages/networkx/drawing/nx_pylab.py:611: Matp
  if cb.is_numlike(alpha):
```


What
is
the
origin
of
clinical
pharmacy
?
Clinical
pharmacists
are
now
an
integral
part
of
the
interdisciplinar
approach
to
patient
care
.
0.36413014
0.3480291
0.533093
0.58236724
0.73318905
0.25139537
0.24161343
0.73318905
0.9104898
0.49427462
0.3316969
0.29641882
0.41309536

```
In [148]: Alignement_graph_bipartite(" When was Waruhiu Itote captured? "," The capture of Dedan

/media/ilan/DATA/Maitrise/env/lib/python3.6/site-packages/networkx/drawing/nx_pylab.py:611: Matp
  if cb.is_numlike(alpha):
```


When
was
Waruhiu
Itote
captured
?
The
capture
of
Dedan
Kimathi
on
21
October
1956
in
Nyeri
signified
the
ultimate
defeat
of
the
Mau
Mau
and
essentially
ended
the
military
offensive
.
0.32229105
0.3119039
0.48711544
0.48711544
0.43305597
0.2523932
0.26740718
0.81501806
0.4101818
0.27140843
0.41309536