

## Task 2: KG Embedding Model

In [1]:

```
%capture  
!pip install ampligraph;
```

In [2]:

```
%tensorflow_version 1.x  
import numpy as np  
import pandas as pd  
import ampligraph  
  
ampligraph.__version__
```

TensorFlow 1.x selected.

Out[2]:

```
'1.3.2'
```

In [3]:

```
import requests  
from ampligraph.datasets import load_from_csv  
  
url = 'https://ampgraphenc.s3-eu-west-1.amazonaws.com/datasets/freebase-237-merged-and-remapped.cs'  
open('freebase.csv', 'wb').write(requests.get(url).content)  
X = load_from_csv('.', 'freebase.csv', sep=',')  
X[:5, ]
```

Out[3]:

```
array([[ 'queens college', 'city university of new york',  
        '/education/educational_institution/students_graduates./education/education/student',  
        'carol leifer'],  
       ['digital equipment corporation',  
        '/business/business_operation/industry', 'computer hardware'],  
       ['/m/0drtv8',  
        '/award/award_ceremony/awards_presented./award/award_honor/award_winner',  
        'laurence mark'],  
       ['the departed',  
        '/award/award_winning_work/awards_won./award/award_honor/award_winner',  
        'leonardo dicaprio'],  
       ['marilyn manson', '/people/person/profession', 'actor']],  
      dtype=object)
```

In [4]:

```
positives_filter = X
```

In [5]:

```
entities = np.unique(np.concatenate([X[:, 0], X[:, 2]]))  
entities
```

Out[5]:

```
array(['/m/011xg5', '/m/011yd2', '/m/011yxg', ..., 'zoology', 'zurich',  
      'zz top'], dtype=object)
```

In [6]:

```
relations = np.unique(X[:, 1])  
relations
```

Out[6]:

```
array(['/american_football/football_team/current_roster./sports/sports_team_roster/position',  
      '/award/award_category/category_of',  
      '/award/award_category/disciplines_or_subjects',  
      '/award/award_category/nominees./award/award_nomination/nominated_for',  
      '/award/award_category/winners./award/award_honor/award_winner',  
      '/award/award_category/winners./award/award_honor/ceremony',  
      '/award/award_ceremony/awards_presented./award/award_honor/award_winner',  
      '/award/award_ceremony/awards_presented./award/award_honor/honored_for',  
      '/award/award_nominated_work/award_nominations./award/award_nomination/nominated_for',  
      ...]
```

'/award/award\_nominee/award\_nominations./award/award\_nomination/award',  
'/award/award\_nominee/award\_nominations./award/award\_nomination/award\_nominee',  
'/award/award\_nominee/award\_nominations./award/award\_nomination/nominated\_for',  
'/award/award\_winner/awards\_won./award/award\_honor/award\_winner',  
'/award/award\_winning\_work/awards\_won./award/award\_honor/award',  
'/award/award\_winning\_work/awards\_won./award/award\_honor/award\_winner',  
'/award/award\_winning\_work/awards\_won./award/award\_honor/honored\_for',  
'/award/hall\_of\_fame/inductees./award/hall\_of\_fame\_induction/inductee',  
'/award/ranked\_item/appears\_in\_ranked\_lists./award/ranking/list',  
'/base/aareas/schema/administrative\_area/administrative\_area\_type',  
'/base/aareas/schema/administrative\_area/administrative\_parent',  
'/base/aareas/schema/administrative\_area/capital',  
'/base/americancoed/celebrity\_impressionist/celebrities\_impersonated',  
'/base/biblioness/bibs\_location/country',  
'/base/biblioness/bibs\_location/state',  
'/base/culturalevent/event/entity\_involved',  
'/base/eating/practicer\_of\_diet/diet',  
'/base/localfood/seasonal\_month/produce\_available./base/localfood/produce\_availability/seas  
onal\_months',  
'/base/locations/continents/countries\_within',  
'/base/marchmadness/ncaa\_basketball\_tournament/seeds./base/marchmadness/ncaa\_tournament\_see  
d/team',  
'/base/petbreeds/city\_with\_dogs/top\_breeds./base/petbreeds/dog\_city\_relationship/dog\_bree  
d',  
'/base/popstra/celebrity/breakup./base/popstra/breakup/participant',  
'/base/popstra/celebrity/canoodled./base/popstra/canoodled/participant',  
'/base/popstra/celebrity/dated./base/popstra/dated/participant',  
'/base/popstra/celebrity/friendship./base/popstra/friendship/participant',  
'/base/popstra/location/vacationers./base/popstra/vacation\_choice/vacationer',  
'/base/saturdaynightlive/snl\_cast\_member/seasons./base/saturdaynightlive/snl\_season\_tenure/  
cast\_members',  
'/base/schemastaging/organization\_extra/phone\_number./base/schemastaging/phone\_sandbox/cont  
act\_category',  
'/base/schemastaging/organization\_extra/phone\_number./base/schemastaging/phone\_sandbox/serv  
ice\_language',  
'/base/schemastaging/organization\_extra/phone\_number./base/schemastaging/phone\_sandbox/serv  
ice\_location',  
'/base/schemastaging/person\_extra/net\_worth./measurement\_unit/dated\_money\_value/currency',  
'/base/x2010fifaworldcupsouthafrica/world\_cup\_squad/current\_world\_cup\_squad./base/x2010fifa  
worldcupsouthafrica/current\_world\_cup\_squad/current\_club',  
'/baseball/baseball\_team/team\_stats./baseball/baseball\_team\_stats/season',  
'/broadcast/content/artist',  
'/business/business\_operation/assets./measurement\_unit/dated\_money\_value/currency',  
'/business/business\_operation/industry',  
'/business/business\_operation/operating\_income./measurement\_unit/dated\_money\_value/curren  
cy',  
'/business/business\_operation/revenue./measurement\_unit/dated\_money\_value/currency',  
'/business/job\_title/people\_with\_this\_title./business/employment\_tenure/company',  
'/celebrities/celebrity/celebrity\_friends./celebrities/friendship/friend',  
'/celebrities/celebrity/sexual\_relationships./celebrities/romantic\_relationship/celebrity',  
'/common/topic/webpage./common/webpage/category',  
'/dataworld/gardening\_hint/split\_to',  
'/education/educational\_degree/people\_with\_this\_degree./education/education/institution',  
'/education/educational\_degree/people\_with\_this\_degree./education/education/major\_field\_of\_  
study',  
'/education/educational\_degree/people\_with\_this\_degree./education/education/student',  
'/education/educational\_institution/campuses',  
'/education/educational\_institution/colors',  
'/education/educational\_institution/school\_type',  
'/education/educational\_institution/students\_graduates./education/education/major\_field\_of\_  
study',  
'/education/educational\_institution/students\_graduates./education/education/student',  
'/education/educational\_institution\_campus/educational\_institution',  
'/education/field\_of\_study/students\_majoring./education/education/major\_field\_of\_study',  
'/education/field\_of\_study/students\_majoring./education/education/student',  
'/education/university/domestic\_tuition./measurement\_unit/dated\_money\_value/currency',  
'/education/university/fraternities\_and\_sororities',  
'/education/university/international\_tuition./measurement\_unit/dated\_money\_value/currency',  
'/education/university/local\_tuition./measurement\_unit/dated\_money\_value/currency',  
'/film/actor/dubbing\_performances./film/dubbing\_performance/language',

```
'/film/actor/film./film/performance/film',
'/film/actor/film./film/performance/special_performance_type',
'/film/director/film', '/film/film/cinematography',
'/film/film/costume_design_by', '/film/film/country',
'/film/film/distributors./film/film_distributor_relationship/film_distribution_medium',
'/film/film/distributors./film/film_distributor_relationship/region',
'/film/film/dubbing.Performances./film/dubbing_performance/actor',
'/film/film/edited_by',
'/film/film/estimated_budget./measurement_unit/dated_money_value/currency',
'/film/film/executive_produced_by',
'/film/film/featured_film_locations',
'/film/film/film_art_direction_by', '/film/film/film_festivals',
'/film/film/film_format', '/film/film/film_production_design_by',
'/film/film/genre', '/film/film/language', '/film/film/music',
'/film/film/other_crew./film/film_crew_gig/crewmember',
'/film/film/other_crew./film/film_crew_gig/film_crew_role',
'/film/film/personal_appearances./film/personal_film_appearance/person',
'/film/film/prequel', '/film/film/produced_by',
'/film/film/production_companies',
'/film/film/release_date_s./film/filmRegional_release_date/filmRegional_debut_venue',
'/film/film/release_date_s./film/filmRegional_release_date/film_release_distribution_medium',
'/film/film/release_date_s./film/filmRegional_release_date/film_release_region',
'/film/film/runtime./film/film_cut/film_release_region',
'/film/film/story_by', '/film/film/written_by',
'/film/film_distributor/films_distributed./film/film_distributor_relationship/film',
'/film/film_set_designer/film_sets_designed',
'/film/film_subject/films',
'/film/person_or_entity_appearing_in_film/films./film/personal_film_appearance/type_of_appearance',
'/film/special_film_performance_type/film_performance_type./film/performance/film',
'/food/food/nutrients./food/nutrition_fact/nutrient',
'/government/government_office_category/officeholders./government/government_position_held/jurisdiction_of_office',
'/government/governmental_body/members./government/government_position_held/legislative_sessions',
'/government/legislative_session/members./government/government_position_held/district_represented',
'/government/legislative_session/members./government/government_position_held/legislative_sessions',
'/government/political_party/politicians_in_this_party./government/political_party_tenure/politician',
'/government/politician/government_positions_held./government/government_position_held/basic_title',
'/government/politician/government_positions_held./government/government_position_held/jurisdiction_of_office',
'/government/politician/government_positions_held./government/government_position_held/legislative_sessions',
'/ice_hockey/hockey_team/current_roster./sports/sports_team_roster/position',
'/influence/influence_node/influenced_by',
'/influence/influence_node/peers./influence/peer_relationship/peers',
'/language/human_language/countries_spoken_in',
'/location/administrative_division/country',
'/location/administrative_division/first_level_division_of',
'/location/capital_of_administrative_division/capital_of./location/administrative_division_capital_relationship/administrative_division',
'/location/country/capital',
'/location/country/form_of_government',
'/location/country/official_language',
'/location/country/second_level_divisions',
'/location/hud_county_place/county',
'/location/hud_county_place/place',
'/location/hud_foreclosure_area/estimated_number_of_mortgages./measurement_unit/dated_integrator/source',
'/location/location/adjoin_s./location/adjoining_relationship/adjoins',
'/location/location/contains',
'/location/location/partially_contains',
'/location/location/time_zones',
'/location/statistical_region/gdp_nominal./measurement_unit/dated_money_value/currency',
```

'/location/statistical\_region/gdp\_nominal\_per\_capita./measurement\_unit/dated\_money\_value/currency',  
  '/location/statistical\_region/gdp\_real./measurement\_unit/adjusted\_money\_value/adjustment\_currency',  
  '/location/statistical\_region/gni\_per\_capita\_in\_ppp\_dollars./measurement\_unit/dated\_money\_value/currency',  
  '/location/statistical\_region/places\_exported\_to./location/imports\_and\_exports/exported\_to',  
  '/location/statistical\_region/religions./location/religion\_percentage/religion',  
  '/location/statistical\_region/rent50\_2./measurement\_unit/dated\_money\_value/currency',  
  '/location/us\_county/county\_seat',  
  '/media\_common/netflix\_genre/titles',  
  '/medicine/disease/notable\_people\_with\_this\_condition',  
  '/medicine/disease/risk\_factors', '/medicine/symptom/symptom\_of',  
  '/military/military\_combatant/military\_conflicts./military/military\_combatant\_group/combatants',  
  '/military/military\_conflict/combatants./military/military\_combatant\_group/combatants',  
  '/music/artist/contribution./music/recording\_contribution/performance\_role',  
  '/music/artist/origin',  
  '/music/artist/track\_contributions./music/track\_contribution/role',  
  '/music/genre/artists', '/music/genre/parent\_genre',  
  '/music/group\_member/membership./music/group\_membership/group',  
  '/music/group\_member/membership./music/group\_membership/role',  
  '/music/instrument/family', '/music/instrument/instrumentalists',  
  '/music/performance\_role/guest\_performances./music/recording\_contribution/performance\_role',  
  '/music/performance\_role/regular\_performances./music/group\_membership/group',  
  '/music/performance\_role/regular\_performances./music/group\_membership/role',  
  '/music/performance\_role/track\_performances./music/track\_contribution/role',  
  '/music/record\_label/artist',  
  '/olympics/olympic\_games/medals\_awarded./olympics/olympic\_medal\_honor/medal',  
  '/olympics/olympic\_games/participating\_countries',  
  '/olympics/olympic\_games/sports',  
  '/olympics/olympic\_participating\_country/athletes./olympics/olympic\_athlete\_affiliation/olympics',  
  '/olympics/olympic\_participating\_country/medals\_won./olympics/olympic\_medal\_honor/medal',  
  '/olympics/olympic\_participating\_country/medals\_won./olympics/olympic\_medal\_honor/olympics',  
  '/olympics/olympic\_sport/athletes./olympics/olympic\_athlete\_affiliation/country',  
  '/olympics/olympic\_sport/athletes./olympics/olympic\_athlete\_affiliation/olympics',  
  '/organization/endowed\_organization/endowment./measurement\_unit/dated\_money\_value/currency',  
  '/organization/non\_profit\_organization/registered\_with./organization/non\_profit\_registration/registering\_agency',  
  '/organization/organization/child./organization/organization\_relationship/child',  
  '/organization/organization/headquarters./location mailing\_address/citytown',  
  '/organization/organization/headquarters./location mailing\_address/country',  
  '/organization/organization/headquarters./location mailing\_address/state\_province\_region',  
  '/organization/organization/place\_founded',  
  '/organization/organization\_founder/organizations Founded',  
  '/organization/organization\_member/member\_of./organization/organization\_membership/organization',  
  '/organization/role/leaders./organization/leadership/organization',  
  '/people/cause\_of\_death/people',  
  '/people/deceased\_person/place\_of\_burial',  
  '/people/deceased\_person/place\_of\_death',  
  '/people/ethnicity/geographic\_distribution',  
  '/people/ethnicity/languages\_spoken', '/people/ethnicity/people',  
  '/people/marriage\_union\_type/unions\_of\_this\_type./people/marriage/location\_of\_ceremony',  
  '/people/person/employment\_history./business/employment\_tenure/company',  
  '/people/person/gender', '/people/person/languages',  
  '/people/person/nationality', '/people/person/place\_of\_birth',  
  '/people/person/places\_lived./people/place\_lived/location',  
  '/people/person/profession', '/people/person/religion',  
  '/people/person/sibling\_s./people/sibling\_relationship/sibling',  
  '/people/person/spouse\_s./people/marriage/location\_of\_ceremony',  
  '/people/person/spouse\_s./people/marriage/spouse',  
  '/people/person/spouse\_s./people/marriage/type\_of\_union',  
  '/people/profession/specialization\_of',  
  '/soccer/football\_player/current\_team./sports/sports\_team\_roster/team',

```
'/soccer/football_team/current_roster./soccer/football_roster_position/position',
'/soccer/football_team/current_roster./sports/sports_team_roster/position',
'/sports/pro_athlete/teams./sports/sports_team_roster/team',
'/sports/professional_sports_team/draft_picks./sports/sports_league_draft_pick/draft',
'/sports/professional_sports_team/draft_picks./sports/sports_league_draft_pick/school',
'/sports/sport/pro_athletes./sports/pro_sports_played/athlete',
'/sports/sports_league/teams./sports/sports_league_participation/team',
'/sports/sports_league_draft/picks./sports/sports_league_draft_pick/school',
'/sports/sports_position/players./american_football/football_historical_roster_position/pos
ition_s',
'/sports/sports_position/players./sports/sports_team_roster/position',
'/sports/sports_position/players./sports/sports_team_roster/team',
'/sports/sports_team/colors',
'/sports/sports_team/roster./american_football/football_historical_roster_position/position
_s',
'/sports/sports_team/roster./american_football/football_roster_position/position',
'/sports/sports_team/roster./baseball/baseball_roster_position/position',
'/sports/sports_team/roster./basketball/basketball_roster_position/position',
'/sports/sports_team/sport', '/sports/sports_team_location/teams',
'/time/event/instance_of_recurring_event', '/time/event/locations',
'/travel/travel_destination/climate./travel/travel_destination_monthly_climate/month',
'/travel/travel_destination/how_to_get_here./travel/transportation/mode_of_transportation',
'/tv/non_character_role/tv_regular_personal_appearances./tv/tv_regular_personal_appearance/
person',
'/tv/tv_network/programs./tv/tv_network_duration/program',
'/tv/tv_personality/tv_regular_personal_appearance./tv/tv_regular_personal_appearance/program',
'/tv/tv_producer/programs_produced./tv/tv_producer_term/producer_type',
'/tv/tv_producer/programs_produced./tv/tv_producer_term/program',
'/tv/tv_program/country_of_origin', '/tv/tv_program/genre',
'/tv/tv_program/languages', '/tv/tv_program/program_creator',
'/tv/tv_program/regular_cast./tv/regular_tv_appearance/actor',
'/tv/tv_program/tv_producer./tv/tv_producer_term/producer_type',
'/tv/tv_writer/tv_programs./tv/tv_program_writer_relationship/tv_program',
'/user/alexander/philosophy/philosopher/interests',
'/user/jg/default_domain/olympic_games/sports',
'/user/ktrueman/default_domain/international_organization/member_states',
'/user/tsegaran/random/taxonomy_subject/entry./user/tsegaran/random/taxonomy_entry/taxonom
y'],
dtype=object)
```

```
In [7]: from ampligraph.evaluation import train_test_split_no_unseen
```

```
num_test = 10000

data = {}
data['train'], data['test'] = train_test_split_no_unseen(X, test_size=num_test, seed=0, allow_dupe
l=True)
```

```
In [8]: print('Train set size: ', data['train'].shape)
print('Test set size: ', data['test'].shape)
```

```
Train set size: (298722, 3)
Test set size: (10000, 3)
```

```
In [9]: from ampligraph.latent_features import ComplEx
```

```
In [10]: model = ComplEx(batches_count=100,
                      seed=0,
                      epochs=100,
                      k=150,
                      eta=5,
                      optimizer='adam',
                      optimizer_params={'lr':1e-3},
                      loss='multiclass_nll',
                      regularizer='LP',
```

```
regularizer_params={'p':3, 'lambda':1e-5},  
verbose=True)
```

```
In [11]: model.fit(data['train'], early_stopping = False)
```

```
Average Loss: 0.028910: 100%|██████████| 100/100 [33:34<00:00, 20.14s/epoch]
```

```
In [12]: from ampligraph.latent_features import save_model, restore_model
```

```
In [13]: save_model(model, './best_model.pkl')
```

```
In [14]: del model
```

```
In [15]: model = restore_model('./best_model.pkl')
```

```
In [16]: if model.is_fitted:  
    print('The model is fit!')  
else:  
    print('The model is not fit! Did you skip a step?')
```

```
The model is fit!
```

```
In [17]: from ampligraph.evaluation import evaluate_performance
```

```
In [18]: ranks = evaluate_performance(data['test'],  
                                 model=model,  
                                 filter_triples=positives_filter, # Corruption strategy filter defin  
                                 use_default_protocol=True, # corrupt subj and obj separately while ev  
                                 verbose=True)
```

```
WARNING - DeprecationWarning: use_default_protocol will be removed in future. Please use corrupt_s  
ide argument instead.
```

```
100%|██████████| 10000/10000 [15:07<00:00, 11.01it/s]
```

```
In [19]: from ampligraph.evaluation import mr_score, mrr_score, hits_at_n_score  
  
mrr = mrr_score(ranks)  
print("MRR: %.2f" % (mrr))  
  
hits_10 = hits_at_n_score(ranks, n=10)  
print("Hits@10: %.2f" % (hits_10))  
hits_3 = hits_at_n_score(ranks, n=3)  
print("Hits@3: %.2f" % (hits_3))  
hits_1 = hits_at_n_score(ranks, n=1)  
print("Hits@1: %.2f" % (hits_1))
```

```
MRR: 0.23  
Hits@10: 0.39  
Hits@3: 0.25  
Hits@1: 0.14
```

```
In [20]: X_unseen = data['test'][:10]
```

```
In [21]:
```

```
unseen_filter = np.array(list({tuple(i) for i in np.vstack((positives_filter, X_unseen))}))
```

In [22]:

```
ranks_unseen = evaluate_performance(  
    X_unseen,  
    model=model,  
    filter_triples=unseen_filter, # Corruption strategy filter defined above  
    corrupt_side = 's+o',  
    use_default_protocol=False, # corrupt subj and obj separately while evaluating  
    verbose=True  
)
```

100%|██████████| 10/10 [00:00<00:00, 10.39it/s]

In [23]:

```
scores = model.predict(X_unseen)
```

In [24]:

```
from scipy.special import expit  
probs = expit(scores)
```

In [25]:

```
pd.DataFrame(list(zip([' '.join(x) for x in X_unseen],  
                      ranks_unseen,  
                      np.squeeze(scores),  
                      np.squeeze(probs))),  
            columns=['statement', 'rank', 'score', 'prob']).sort_values("score")
```

Out[25]:

	statement	rank	score	prob
7	chris nurse /people/person/place_of_birth hamm...	4482	3.299764	0.964421
2	kymx /broadcast/content/artist richard marx	457	9.575715	0.999931
8	benjamin bratt /award/award_nominee/award_nomi...	58	10.927811	0.999982
5	david cross /influence/influence_node/influenc...	128	14.479103	1.000000
3	morehouse college /business/business_operation...	13	14.723392	1.000000
0	holly hunter /award/award_nominee/award_nomina...	48	16.066914	1.000000
6	norman stiles /award/award_nominee/award_nomin...	4	16.827826	1.000000
4	erika christensen /award/award_nominee/award_n...	46	17.347830	1.000000
1	pop music /music/genre/artists alanis morissette	26	17.611797	1.000000
9	tyler perry /award/award_nominee/award_nominat...	45	18.875919	1.000000

In [25]: