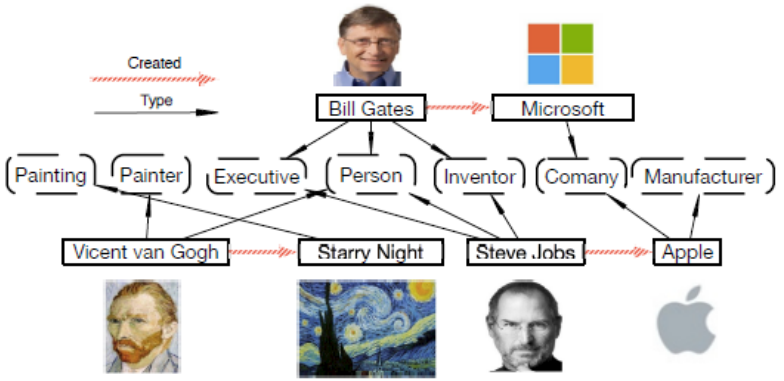
Entity set expansion in knowledge graph: a heterogeneous information network perspective

Chuan SHI, Jiayu DING, Xiaohuan CAO, Linmei HU, Bin WU, Xiaoli LI

Frontiers of Computer Science, DOI: 10.1007/s11704-020-9240-8

Problems & Ideas

- Problems of Entity Set Expansion in Knowledge Graph.
  - It costs huge time and space to discover meta paths in KG.
  - Objects connected by a relation may belong to multiple types, which will cause ambiguity.
  - After extracting meta paths between seeds, there are only limited positive samples and no negative samples for weight learning.
  - Ignore the text information of entities in Wikipedia.
- Ideas:
  - We propose CoMeSE++ which combines structural information revealed by a KG and textual information in Wikipedia for ESE
  - We propose MuTyPaths which are fine-grained and can more accurately represent the inherent relation among seed entities avoiding generating too many false positives for ESE.



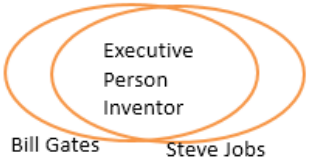
An example with ambiguity in DBpedia

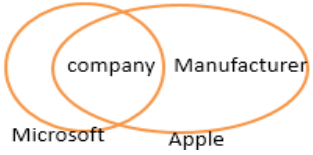
Meta Path: $Obj \xrightarrow{Created} Obj$

Instances:

Bill Gates $\xrightarrow{Created}$ Microsoft

Steve Jobs $\xrightarrow{Created}$ Apple





MuTyPath : $(Person \wedge Inventor \wedge Executive) \xrightarrow{Created} (Company)$

The average performance of different methods

Methods		Link	Neighbor	DeepWalk	PCRW-2	PCRW-3	PCRW-4	MP_ESE	Text-Only	CoMeSE_IIc	CoMeSE_PU	CoMeSE++
Company	MAP	0.118	0.180	0.112	0.167	0.373	0.64	0.400	0.885	0.878	0.950	**0.956**
	P@10	0.050	0.070	0.025	0.134	0.345	0.530	0.500	0.88	0.825	0.950	**0.960**
	P@30	0.012	0.020	0.01	0.070	0.208	0.390	0.400	0.465	0.826	0.880	**0.908**
	P@60	0.093	0.001	0.01	0.044	0.167	0.302	0.300	0.331	0.800	0.805	**0.815**
Writer	MAP	0.035	0.648	0.073	0.690	0.737	0.770	0.870	0.992	0.882	0.999	**0.999**
	P@10	0.010	0.605	0.020	0.659	0.690	0.707	0.800	0.500	0.820	0.995	**0.995**
	P@30	0.015	0.453	0.008	0.625	0.640	0.701	0.740	0.125	0.740	0.956	**0.960**
	P@60	0.028	0.240	0.008	0.370	0.380	0.440	0.460	0.085	0.460	0.832	**0.837**
Actor	MAP	0.236	0.125	0.004	0.010	0.329	0.504	0.600	0.894	0.779	0.930	**0.956**
	P@10	0.060	0.100	0.000	0.000	0.215	0.325	0.500	0.885	0.750	0.910	**0.935**
	P@30	0.058	0.033	0.003	0.000	0.215	0.315	0.500	0.522	0.730	0.880	**0.883**
	P@60	0.054	0.016	0.004	0.020	0.176	0.315	0.520	0.36	0.730	0.850	**0.853**
Movie	MAP	0.004	0.214	0.098	0.290	0.330	0.620	0.780	0.750	0.790	0.922	**0.938**
	P@10	0.000	0.150	0.055	0.230	0.250	0.530	0.780	0.720	0.750	0.910	**0.965**
	P@30	0.000	0.070	0.063	0.120	0.215	0.350	0.710	0.380	0.690	0.825	**0.842**
	P@60	0.002	0.045	0.062	0.070	0.176	0.270	0.620	0.300	0.650	0.721	**0.733**