TransNet: Translation-Based Network Representation Learning for Social Relation Extraction

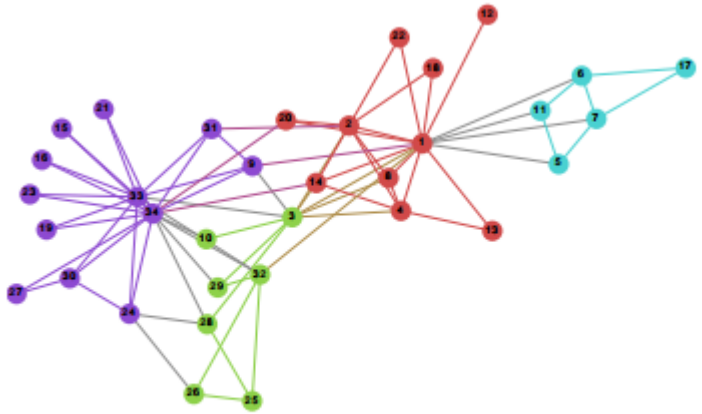
Cunchao Tu

NLP Group, Dept. of CS&T, Tsinghua University

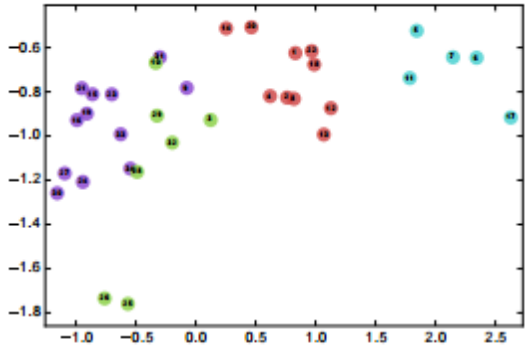
IJCAI 2017

Network Representation Learning

- Learn a low-dimensional representation vector for each vertex in networks



(a) Input: Karate Graph



(b) Output: Representation

Network Representation Learning

- Representations as features
  - Vertex classification -> User profiling/ Anomaly detection

Network Representation Learning

- Representations as features
  - Vertex classification -> User profiling/ Anomaly detection
  - Link prediction -> Friend recommendation

Challenges

- Typical NRL models
  - simplify each edge as a binary or continuous value.
  - ignore rich semantic information on edges.
- Typical NRL tasks
  - Vertex classification
  - Link prediction

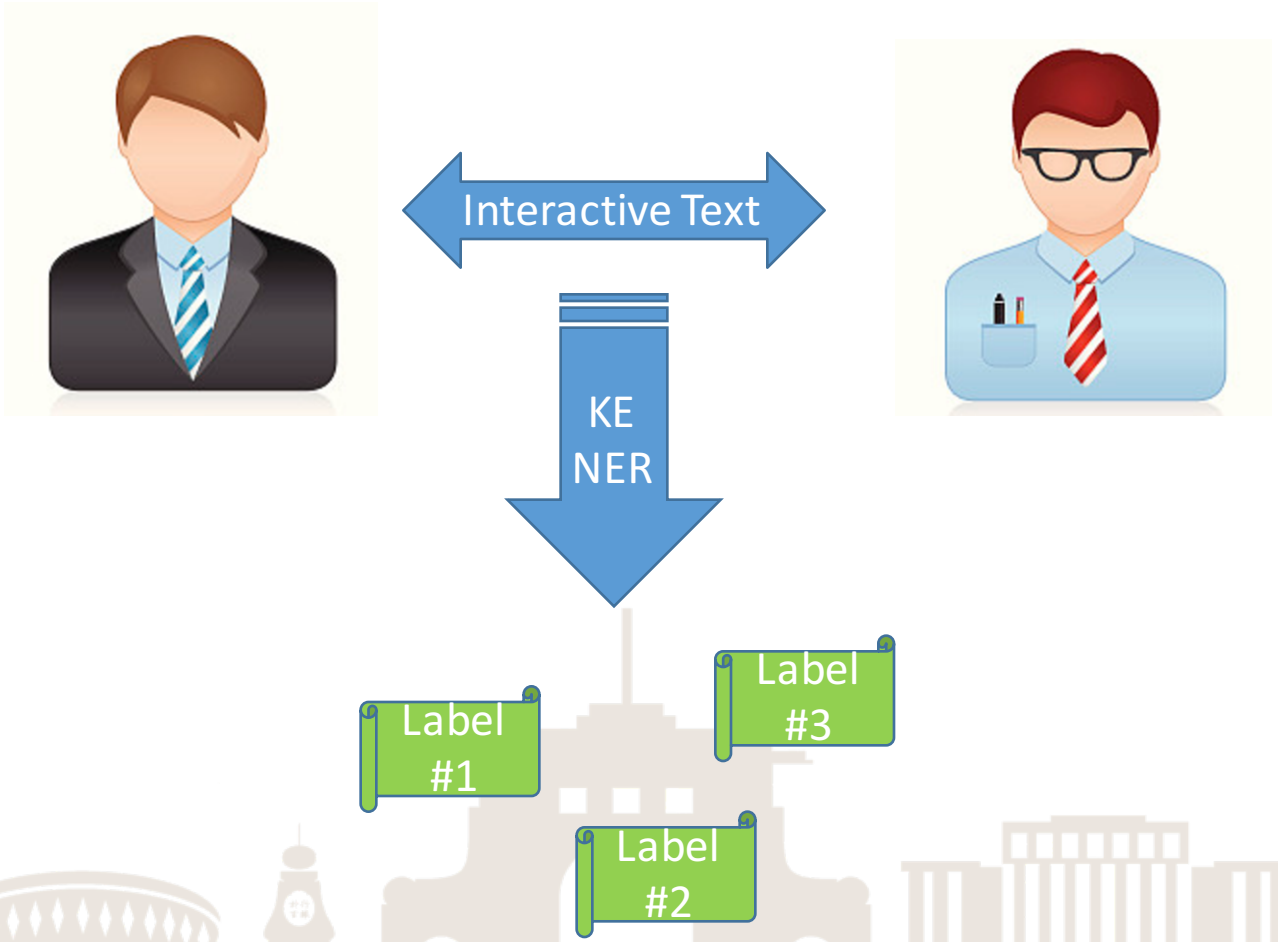
Can we model and predict detailed relations between social network vertices?

A New Task – Social Relation Extraction

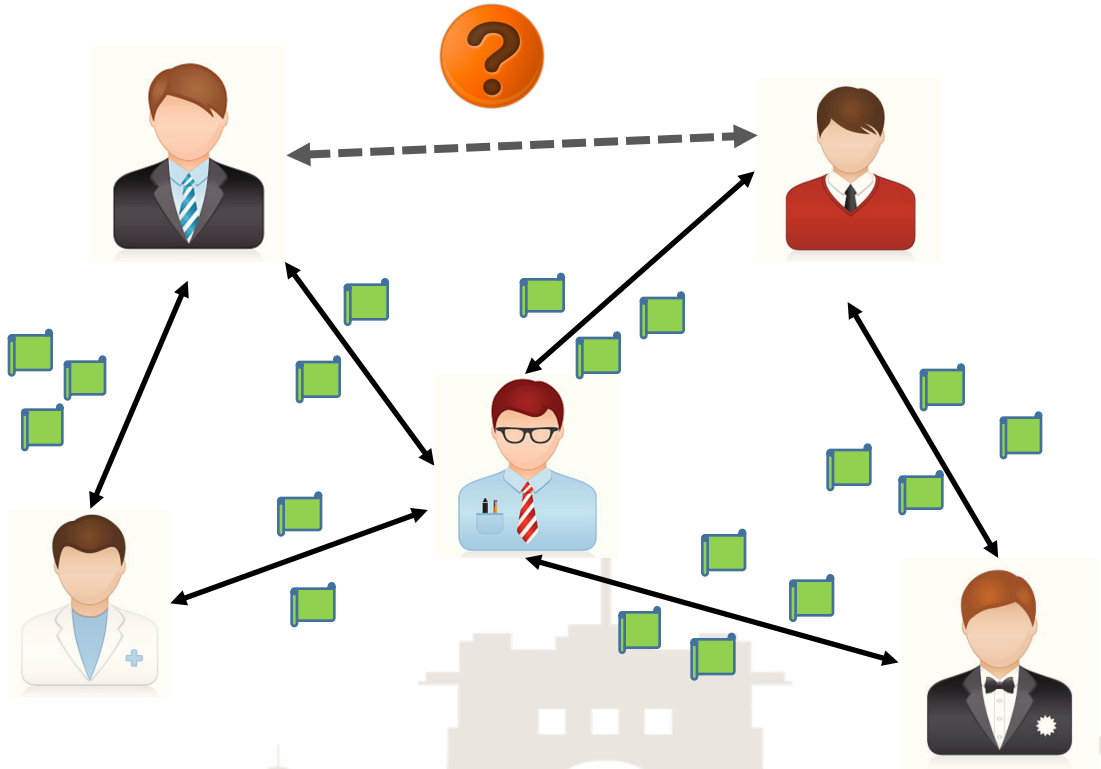
- Extract detailed relations between social network vertices
  - A new way to evaluate relation modeling of NRL models
- Relation Extraction/Prediction in knowledge graphs
  - Relation categories are well pre-defined
  - Relational facts are annotated precisely with human efforts

Definitions of SRE

- Relations

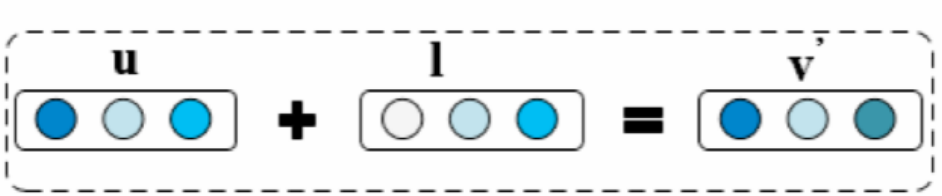


Definitions of SRE



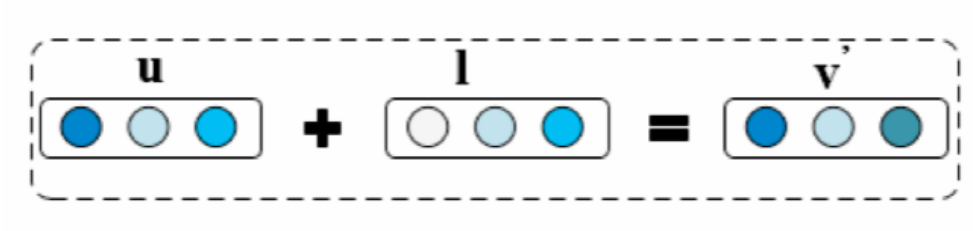
TransNet: Translation-Based NRL

- Translation Mechanism
  - “Interest” vector **u, v**
  - “Skill” vector **u’, v’**



TransNet: Translation-Based NRL

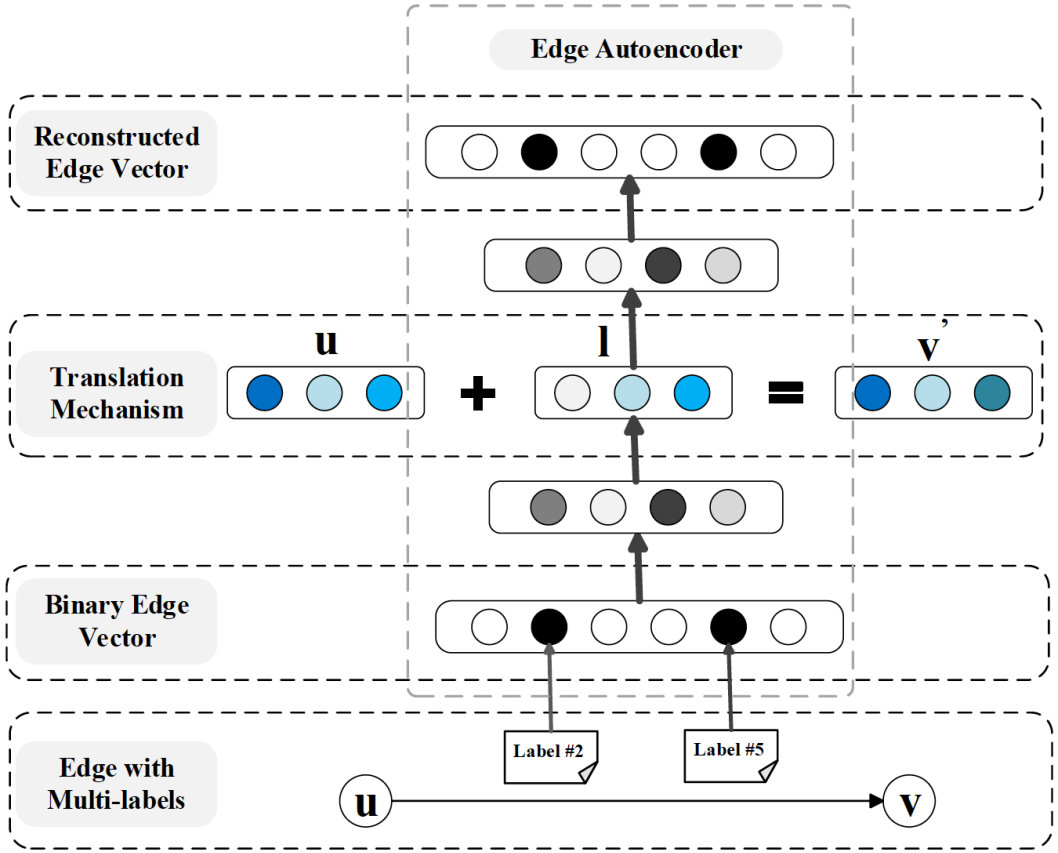
- Translation Mechanism



$$\mathcal{L}_{trans} = \max(\gamma + d(\mathbf{u} + \mathbf{l}, \mathbf{v}') - d(\hat{\mathbf{u}} + \hat{\mathbf{l}}, \hat{\mathbf{v}}'), 0)$$

TransNet: Translation-Based NRL

- Edge Representation Construction



Experiments

- Datasets: Arnetminer

Datasets	**Arnet-S**	**Arnet-M**	**Arnet-L**
Vertices	187,939	268,037	945,589
Edges	1,619,278	2,747,386	5,056,050
Train	1,579,278	2,147,386	3,856,050
Test	20,000	300,000	600,000
Valid	20,000	300,000	600,000
Labels	100	500	500
ML Proportion (%)	42.46	63.74	61.68

- Evaluation Metrics: Hits@k, MeanRank

Experiments

- SRE
  - Typical NRL models
  - TransE and TransNet

Table 2: SRE results on Arnet-S. ($\times 100$ for $hits@k$, $\alpha = 0.5$ and $\beta = 20$)

Metric	$hits@1$	$hits@5$	$hits@10$	MeanRank	$hits@1$	$hits@5$	$hits@10$	MeanRank
DeepWalk	13.88	36.80	50.57	19.69	18.78	39.62	52.55	18.76
LINE	11.30	31.70	44.51	23.49	15.33	33.96	46.04	22.54
node2vec	13.63	36.60	50.27	19.87	18.38	39.41	52.22	18.92
TransE	39.16	78.48	88.54	5.39	57.48	84.06	90.60	4.44
TransNet	**47.67**	**86.54**	**92.27**	**5.04**	**77.22**	**90.46**	**93.41**	**4.09**

Table 4: SRE results on Arnet-L. ($\times 100$ for $hits@k$, $\alpha = 0.5$ and $\beta = 50$)

Metric	$hits@1$	$hits@5$	$hits@10$	MeanRank	$hits@1$	$hits@5$	$hits@10$	MeanRank
DeepWalk	5.41	16.17	23.33	102.83	7.59	17.71	24.58	100.82
LINE	4.28	13.44	19.85	114.95	6.00	14.60	20.86	112.93
node2vec	5.39	16.23	23.47	102.01	7.53	17.76	24.71	100.00
TransE	15.38	41.87	55.54	32.65	23.24	47.07	59.33	30.64
TransNet	**28.85**	**66.15**	**75.55**	**29.60**	**56.82**	**73.42**	**78.60**	**27.40**

Experiments

- Label comparison

Table 5: Label comparisons on Arnet-S. ($\times 100$ for $hits@k$)

Tags	Top 5 labels				Bottom 5 labels			
Metric	$hits@1$	$hits@5$	$hits@10$	MeanRank	$hits@1$	$hits@5$	$hits@10$	MeanRank
TransE	58.82	85.68	91.61	**3.70**	52.21	82.03	87.75	5.65
TransNet	**77.26**	**90.35**	**93.53**	3.89	**78.27**	**90.44**	**93.30**	**4.18**

Experiments

- Case study: “A. Swami”,”

Table 6: Recommended top-3 labels for each neighbor.

Neighbors	TransE	TransNet
Matthew Duggan	**ad hoc network**; wireless sensor network; wireless sensor networks	**management system**; **ad hoc network**; wireless sensor
K. Pelechrinis	**wireless network**; wireless networks; ad hoc network	**wireless network**; wireless sensor network; **routing protocol**
Oleg Korobkin	**wireless network**; wireless networks; **wireless communication**	**resource management**; **system design**; **wireless network**

Conclusion

- A new task – Social Relation Extraction
  - Relations -> Automatically constructed label set
- Translation-based NRL – TransNet
  - Translation mechanism
  - Deep autoencoder
- 10%~20% improvements than TransE on SRE

Q&A

Homepage

Code and Datasets

Email: tucunchao@gmail.com