

# Demographic Inference on Twitter using Recursive Neural Networks

Sunghwan Mac Kim, Qionghai Xu, Lizhen Qu, Stephen Wan and Cécile Paris

Data61, CSIRO, Australia

{Mac.Kim, Qionghai.Xu, Lizhen.Qu, Stephen.Wan, Cecile.Paris}@data61.csiro.au

## Abstract

In social media, demographic inference is a critical task in order to gain a better understanding of a cohort and to facilitate interacting with one's audience. Most previous work has made independence assumptions over topological, textual and label information on social networks. In this work, we employ recursive neural networks to break down these independence assumptions to obtain inference about demographic characteristics on Twitter. We show that our model performs better than existing models including the state-of-the-art.

## 1 Introduction

Social media is a popular public platform for communicating, sharing information and expressing opinions. Millions of users discuss a variety of topics such as politics or sports. Valuable insights can be obtained by analysing social media content (e.g., mining consumer preferences), and, consequently, social media data is now a valuable resource. Accordingly, social media analytics have received much attention among researchers and companies (Wan and Paris, 2014; Valdes et al., 2015; Zubiaga et al., 2016).

Inferring demographic characteristics from social media is a useful mechanism to gain a better understanding of a cohort and one's audience, and to facilitate interacting with that audience. Many researchers have studied ways to infer demographic attributes of Twitter users, such as age (Mislove et al., 2011; Mohammady Ardehaly and Culotta, 2015), gender (Filippova, 2012; Taniguchi et al., 2015), occupation (Preoțiuc-Pietro et al., 2015; Kim et al., 2016), location (Jurgens et al., 2015; Jayasinghe et al., 2016) or politi-

cal preferences (Volkova et al., 2014; McCormick et al., 2015).

A common approach to infer demographic characteristics is the use of supervised classifiers trained on textual features. The main limitation of this approach is that it makes little use of the network topology. Several network embedding methods have been proposed to learn distributed dense representations for vertices in graphs: DeepWalk (Perozzi et al., 2014) or LINE (Tang et al., 2015). While these two models can capture the topological structure of social networks, their performances are still limited, as the text features associated with vertices are not considered. For instance, text messages that Twitter users post, tweets, can offer great potential to enhance the vertex embeddings. Yang et al. (2015) proposed the Text-Associated DeepWalk (TADW) to enhance the discriminative power of the vertex embeddings by incorporating the text information into the embedding generation process. Although this matrix factorisation framework is effective on the vertex classification task, it can produce suboptimal embeddings. This is because the label information is not exploited in the unsupervised framework. More recently, Pan et al. (2016) proposed the Tri-Party Deep Network Representation (TriDNR), a method that incorporates the label information in addition to the text and topological information. However, in TriDNR, the text in a vertex is assumed to be independent of neighbour vertices. Furthermore, the two optimisation problems (learning vertex embeddings and training discriminative classifiers) are tackled separately. This is also true of TADW.

In this paper, we tackle the problem of inferring demographic characteristics from social networks as a vertex classification task on graphs. We employ recursive neural networks (RNNs) to infer three demographic attributes of Twitter users

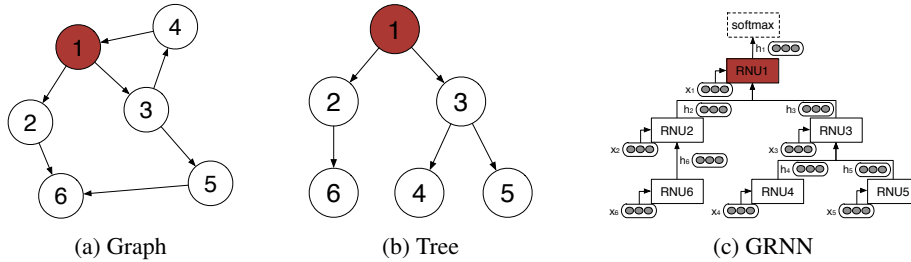


Figure 1: An illustration of GRNN construction steps: (a) The graph consisting of six vertices; (b) The tree converted from the graph using breadth-first search; (c) The GRNN constructed from the tree. The target vertex is marked in red.

(age, gender and user type) based on network topology, text content and label information. Our model breaks down the independence assumption by leveraging RNNs on paths in graphs. We show that our model achieves better performance compared to existing models including the state-of-the-art. While high performance is achieved using solely neural network models, more importantly, we find that different demographic inference tasks benefit to varying the topological size of RNNs. To our knowledge, there has been little previous work applying neural network based methods to the problem of inferring social media demographics.

## 2 Graph Recursive Neural Networks

RNNs are deep learning models that recursively compose the vector of a parent unit from those of child units over a given structure in topological order (Pollack, 1990). They have shown to be very effective for various natural language processing (NLP) tasks, capturing syntactic and semantic composition (Socher et al., 2011; Qian et al., 2015). In this section, we describe the framework of Graph RNNs (GRNNs) (Xu et al., 2017) to classify the vertices of a graph. This framework allows us to infer the demographic characteristics of social media users.

We formally define the problem of Twitter vertex classification as follows: A Twitter social network is defined as  $G = (V, E)$ , where  $V$  is the set of vertices (users) and  $E$  is the set of edges (relationships) between the vertices. Each edge  $e \in E$  is an ordered pair  $e = (v_i, v_j)$ , where  $v_i, v_j \in V$ , that is unweighted<sup>1</sup> ( $w_{ij} = 0$ ) but di-

<sup>1</sup>While the edges in the Twitter social network can be weighted (e.g., based on the importance of relationship), we only take into account an unweighted social network in this study.

rected ( $(i, j) \neq (j, i)$ ). Each  $v_i$  is associated with a pair of  $(x_i, y_i)$ , where  $x_i \in \mathcal{X}$  is a feature vector and  $y_i \in \mathcal{Y}$  is a particular label that depends on  $x_i$ .  $\mathcal{X}$  and  $\mathcal{Y}$  thus denote a set of feature vectors and a set of possible predicted labels in  $G$ , respectively. Our goal is then to predict the most likely label  $\hat{y}_t \in \mathcal{Y}$  for  $v_t \in V$ , which is the target vertex to be classified:  $\hat{y}_t = \operatorname{argmax}_{y_t \in \mathcal{Y}} P_\theta(y_t | v_t, G, \mathcal{X})$  using a RNN with parameters  $\theta$ .

GRNNs contain five components that will be presented in this section: 1) Graph-to-Tree Conversion, 2) Word Embedding layer, 3) Recursive Neural Unit layer and 4) Softmax Output layer, as illustrated in Figure 1.

### 2.1 Graph-to-Tree Conversion

A graph is converted to tree structures before constructing a RNN for each tree. Specifically, we construct a tree  $T = (V_T, E_T)$  of depth  $d$  rooted at  $v_t$  using a breadth-first search algorithm from  $G$ .  $V_T$  and  $E_T$  are the sets of vertices and edges in the tree.  $(v, w) \in E_T$  denotes an edge from a parent vertex  $v$  to a child vertex  $w$ .

### 2.2 Word Embeddings

Let  $S_i = \{w_1, w_2, \dots, w_R\}$  be texts (e.g., tweets) consisting of  $R$  words, which are associated with a vertex  $v_i$ . Every word  $w_r$  is converted into a real-valued vector  $e_r$  by looking up the embedding matrix  $E \in \mathbb{R}^{d_w \times |V|}$ , where  $d_w$  is the size of word embedding and  $|V|$  is a vocabulary size. The matrix  $E$  is initialised using the Skip-gram model (Mikolov et al., 2013).  $S_i$  is then fed into the next layer as a real-valued feature vector  $x_i = \{e_1, e_2, \dots, e_R\}$ .

### 2.3 Recursive Neural Units

Once we construct a tree from a graph, we build a RNN using one of two types of recursive neural units (RNUs) for each vertex  $v_k \in T$ : Naive

Recursive Neural Unit (NRNU) and Long Short-Term Memory Unit (LSTMU).

### 2.3.1 Graph Naive Recursive Neural Net

Each NRNU for a vertex  $v_k$  take its feature vector  $x_k$  and a hidden state  $\tilde{h}_k$  as input. Max pooling produces  $\tilde{h}_k$  from all hidden state vectors  $h_r$  of the child vertices  $v_r$  of  $v_k$ .<sup>2</sup> A hidden state vector  $h_k$  of  $v_k$  is obtained using weight matrices, followed by a non-linear function  $\tanh$ :

$$\begin{aligned}\tilde{h}_k &= \max_{v_r \in C(v_k)} \{h_r\} \\ h_k &= \tanh(W^{(h)}x_k + U^{(h)}\tilde{h}_k + b^{(h)})\end{aligned}$$

where  $C(v_k)$  is the set of child vertices of  $v_k$  ( $v_r \in C(v_k)$ ) and  $h_r$  is a hidden state of a child vertex  $v_r$ .  $W^{(h)}$  and  $U^{(h)}$  are weight matrices, and  $b^{(h)}$  is a bias vector for model parameters. In this paper, we refer to Graph Naive Recursive Neural Network as GNRNN incorporating NRNU as a RNU.

### 2.3.2 Graph Long Short-Term Memory Net

LSTMU (Hochreiter and Schmidhuber, 1997) was originally proposed to tackle a sequential labelling problem and it is able to model long-range dependencies by incorporating gated memory cells. At each time step, LSTMU takes the sequential input vector and the previous hidden state vector to produce the next hidden state. In this study, LSTMU is employed as a RNU to represent a vertex in a tree, and it naturally captures the relationships between vertices.

For a vertex  $v_k$ , LSTMU takes  $x_k$  and  $\tilde{h}_k$  as input, and generates the input, forget and output gate signals, denoted as  $i_k$ ,  $f_k$  and  $o_k$  respectively. It produces a memory cell state  $c_k$  and hidden state  $h_k$  with respect to a vertex  $v_k$ :

$$\begin{aligned}\tilde{h}_k &= \max_{v_r \in C(v_k)} \{h_r\} \\ \tilde{c}_k &= \tanh(W^{(c)}x_k + U^{(c)}\tilde{h}_k + b^{(c)}) \\ i_k &= \sigma(W^{(i)}x_k + U^{(i)}\tilde{h}_k + b^{(i)}) \\ f_{kr} &= \sigma(W^{(f)}x_k + U^{(f)}h_r + b^{(f)}) \\ o_k &= \sigma(W^{(o)}x_k + U^{(o)}\tilde{h}_k + b^{(o)}) \\ c_k &= i_k \odot \tilde{c}_k + \sum_{v_r \in C(v_k)} f_{kr} \odot c_r \\ h_k &= o_k \odot \tanh(c_k)\end{aligned}$$

where  $\odot$  refers to element-wise product and  $\sigma$  indicates the sigmoid function.  $W^{(*)}$ ,  $U^{(*)}$  and  $b^{(*)}$

<sup>2</sup>Our preliminary results demonstrated that the max pooling strategy achieved better performance than sum and average poolings.

Dataset	Gender	Age	UserType
Users $ V $	5,367	6,482	3,017
Relationships $ E $	5,088	6,514	38,785
Labels $ Y $	2	2	3
Avg. degree	1.90	2.01	25.71
Num. texts	383,425	701,889	3,017
Vocab size	12,558	20,946	615

Table 1: Statistics of the three Twitter social networks. Num. texts indicates the number of tweets for Gender and Age, and the number of profile descriptions for UserType.

are LSTM parameters. We call a tree-structured network topology consisting of LSTMUs as Graph Long Short-Term Memory Net (GLSTMN).

### 2.4 Softmax Output

At the end, the hidden state  $h_t$  is fed into a softmax classifier to predict the label  $y_t$  of the target vertex  $v_t$  after calculating the hidden states of all vertices in  $T$ :  $P_\theta(y_t|v_t, G, \mathcal{X}) = \text{softmax}(W^{(s)}h_t + b^{(s)})$   
 $\hat{y}_t = \text{argmax}_{y_t \in \mathcal{Y}} P_\theta(y_t|v_t, G, \mathcal{X})$ .

## 3 Experimental Setup

In this section, we provide an overview of datasets and the models that are evaluated in the experiments.

### 3.1 Datasets

We evaluate the effectiveness of our model on three types of social networks: gender, age and user type classification. Twitter users follow others or are followed, and two types of relationships are used to build the social networks: *friend* and *follower*. A user is associated with others via the following relationship, the user’s *friend* in Twitter’s terminology. *Follower* relationships indicate that a user receives all the tweets from those the user follows.

- **Gender** (Volkova, 2014) is a Twitter social network encoding *friend* relationships between users. The labels of this network are *Male* and *Female*.
- **Age** (Volkova, 2014) is a Twitter social network encoding *friend* relationships between users. The labels of this network are *Young* (18-23 years old) and *Old* (25-30 years old).
- **UserType** (Kim et al., 2017) is a Twitter social network encoding *follower* relationships between users. The labels represent the types of Twitter users: *Individual*, *Organisation* and *other*.

To generate text features of vertices, up to 1K tweets per user are used in Gender and Age, whereas Twitter user profile descriptions are used in UserType. All words are stemmed, and then stop words and words with document frequency less than 10 are removed. The statistics of the datasets are summarised in Table 1.

### 3.2 Evaluated Models

We compare the GRNN model with several existing models to assess vertex classification performance.

- **Lexica (LX)** (Sap et al., 2014): a lexicon-based method produced from Twitter to calculate the scores of gender and age. These scores are used to predict their labels for users.
- **Logistic Regression (LR)** (Hosmer Jr et al., 2013): a commonly used linear model in the NLP community, only using textual contents in vertices. Bag-of-words feature vectors are generated without incorporating any topological information of a network.
- **Label Propagation (LP)** (Wang and Zhang, 2006): a graph-based semi-supervised learning model, where label probabilities are propagated to all unlabelled neighbours. The probability derivation steps are terminated for the remaining vertices when all label probabilities converge.
- **Text-Associated DeepWalk (TADW)** (Yang et al., 2015): an unsupervised vertex embedding learning method. Low-dimensional representations of vertices are induced both from their texts and graph relationships based on inductive matrix factorisation.
- **Tri-Party Deep Network Representation (TriDNR)** (Pan et al., 2016): two neural networks incorporating the texts, relationships and labels of vertices in graphs. As in TADW, unlabelled vertices are classified using Support Vector Machines (SVMs) (Cortes and Vapnik, 1995) trained on learned vertex embeddings.

### 3.3 Experiment Settings

In our experiments, we follow the standard experimental protocol for vertex classification task. More specifically, we evaluate classification ac-

Labelled Nodes	20%	40%	60%	80%
LX	50.72	50.35	53.29	57.37
LR	57.37	59.91	61.97	62.37
LP	55.92	60.53	60.39	61.05
TADW	53.07	50.93	54.04	51.47
TriDNR	50.99	49.56	48.03	56.58
GNNRNN_d0	57.37	61.67	63.03	65.79
GNNRNN_d1	57.76	61.84	63.55	66.05
GNNRNN_d2	57.76	61.84	63.55	66.05
GLSTMN_d0	<b>58.29</b>	<b>63.68</b>	<b>64.21</b>	<b>68.16</b>
GLSTMN_d1	57.37	62.46	63.95	<b>68.16</b>
GLSTMN_d2	57.37	62.46	63.95	<b>68.16</b>

Table 2: Vertex classification results on Gender (e.g., GNNRNN with depth 1 is represented by GNNRNN\_d1). Numbers in bold represent the highest performance in each column in all tables.

Labelled Nodes	20%	40%	60%	80%
LX	50.10	50.69	46.76	47.95
LR	67.75	71.15	72.97	73.70
LP	67.47	71.61	71.86	73.97
TADW	66.81	68.80	69.31	68.89
TriDNR	58.13	57.60	49.66	57.53
GNNRNN_d0	72.46	74.56	74.62	74.79
GNNRNN_d1	72.53	71.61	76.69	75.89
GNNRNN_d2	72.46	71.98	76.14	76.44
GLSTMN_d0	73.49	74.47	75.86	77.53
GLSTMN_d1	<b>73.91</b>	<b>75.02</b>	77.93	<b>80.82</b>
GLSTMN_d2	73.43	74.65	<b>78.48</b>	80.27

Table 3: Vertex classification results on Age.

curacy<sup>3</sup> with different training ratios, increasing from 20% to 80%. For each training ratio, we randomly generate 5 different training datasets. For each training dataset, we run 10 trials and record the highest accuracy on each testing dataset. We then report the average accuracy for the same ratio of training datasets.

We test six different model architectures using NRNU and LSTMU with three different tree depths ( $d = 0, d = 1, d = 2$ ). The tree depth corresponds to the number of hops between users in a social network. For all the GRNN models, the size of the hidden units is set to 200. We use Adam (Duchi et al., 2011) with a batch size of 20 as the optimisation method that automatically adapts the learning rate in training. The initial learning rate is set to 0.1 for LR and LP, and 0.01 for all GRNNs.

## 4 Results and Analysis

In this section, we present the experimental results and analysis on vertex classification for the three

<sup>3</sup>We report accuracy following previous work for Age and Gender datasets, consisting of balanced label distributions. Although UserType is imbalanced, accuracy is reported in this paper for simplicity and consistency. Previous work also reports classification accuracy for this task.

Labelled Nodes	20%	40%	60%	80%
LR	74.76	76.18	76.98	77.91
LP	73.41	74.50	75.99	75.33
TADW	72.62	75.67	76.43	75.69
TriDNR	74.90	77.91	<b>80.37</b>	<b>81.62</b>
GNRNN_d0	77.02	77.96	78.51	79.21
GNRNN_d1	73.97	77.10	78.66	79.37
GNRNN_d2	NA	NA	NA	NA
GLSTMN_d0	<b>77.16</b>	<b>78.63</b>	79.40	79.77
GLSTMN_d1	72.41	76.51	78.53	80.46
GLSTMN_d2	NA	NA	NA	NA

Table 4: Vertex classification results on UserType. NA stands for “Not Available”.

networks. Numbers in bold represent the highest performance in each column in all tables.

As shown in Table 2, GNRNN and GLSTMN perform better than the evaluated models for the task of gender prediction, with no performance gain when the depth of the tree is increased. Namely, GLSTMN\_d0 is the best performing model for this task. Table 3 shows that GLSTMN also achieves the best performance for age prediction. For this task, however, performance increases with the depth of the tree. For the task of user type classification, Table 4 shows that we attain the best performance with tree depth  $d = 1$  for GLSTMN for the training ratio of 80%. Note that we could not train the GRNN models with tree depth 2 (GNRNN\_d2 and GLSTMN\_d2) on the UserType dataset on our Linux server with 64G memory due to the lack of memory. As shown in Table 1, the average degree of UserType is roughly 10 times larger than that of Gender or Age. The degree of a vertex in a graph indicates the number of edges connected to adjacent vertices. It means that the GRNN models could have approximately 625 ( $= 25 \times 25$ ) vertices in average for the tree depth of 2. Our experiments show that a dense graph consisting of high degree vertices is intractable under the GRNN model.

Interestingly, LR and LP are effective methods for Gender and Age compared to TADW and TriDNR. In particular, TADW performs poorly on Gender, and TriDNR marginally outperforms GLSTMN on the dense social network, UserType.

To summarise, we observed four findings:

1. The GRNN models (GNRNN and GLSTMN) overall outperform the existing models by a noticeable margin in most cases (except for UserType), showing the benefit of RNNs for social network inferences on vertices.
2. The LSTM unit in GLSTMN gives superior

performance over the NRN unit in GNRNN regardless of datasets. Our results are in line with other findings, showing that the LSTM works consistently better than the standard recurrent neural network.

3. GRNNs have different optimal tree depths for each demographic inference task. Tree depth does not improve inference performance for Gender, indicating that only text information is sufficient without incorporating network information. For age, tree depth allows GRNNs to be more effective although its increase does not consistently lead to better performance for GLSTMN. Similarly, tree depth increases the performance of GRNNs for UserType. We hypothesise this may be related to the nature of social interactions (e.g., Twitter users in the similar age group are more likely to interact).
4. The matrix factorisation-based methods (TADW and TriDNR) relatively work well on datasets having high degree vertices, whereas the GRNN models achieve relatively good performance on graphs containing low degree vertices.

## 5 Conclusions and Future Work

In this paper we tackled the demographic inference problem on Twitter as vertex classification on a graph using GRNNs, demonstrating their effectiveness against strong models for selected datasets. The RNN framework provides an effective way to incorporate network, text and label information for Twitter demographic inference. However, different demographic inference tasks benefit to varying the tree depth of GRNN models.

As our future work, we plan to employ other state-of-the-art deep learning models (Yang et al., 2016; Kipf and Welling, 2017) that vary in the nature of the dependency between network, text and label information for demographic inference to confirm the effectiveness of our proposals.

## Acknowledgments

The authors are grateful to three anonymous ACL reviewers. We would also like to thank Brian Jin, who ran the data collection for this study.

## References

- Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20(3):273–297.
- John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Machine Learning Research* 12:2121–2159.
- Katja Filippova. 2012. User demographics and language in an implicit social network. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, Jeju Island, Korea, pages 1478–1488.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9(8):1735–1780.
- David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. 2013. *Applied logistic regression*, volume 398. John Wiley & Sons.
- Gaya Jayasinghe, Brian Jin, James Mchugh, Bella Robinson, and Stephen Wan. 2016. CSIRO Data61 at the WNUT geo shared task. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*. The COLING 2016 Organizing Committee, Osaka, Japan, pages 218–226.
- David Jurgens, Tyler Finethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015. Geolocation prediction in Twitter using social networks: A critical analysis and review of current practice. In *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*. pages 188–197.
- Sunghwan Mac Kim, Cécile Paris, Robert Power, and Stephen Wan. 2017. Distinguishing individuals from organisations on Twitter. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Perth, Australia, WWW '17, pages 805–806.
- Sunghwan Mac Kim, Stephen Wan, and Cécile Paris. 2016. Occupational representativeness in Twitter. In *Proceedings of the 21st Australasian Document Computing Symposium, ADCS 2016, Caulfield, VIC, Australia, December 5-7, 2016*. pages 57–64.
- Thomas N Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Tyler H. McCormick, Hedwig Lee, Nina Cesare, Ali Shojaie, and Emma S. Spiro. 2015. Using Twitter for demographic and social science research: Tools for data collection and processing. *Sociological Methods & Research*.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR* abs/1301.3781.
- Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist. 2011. Understanding the demographics of Twitter users. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, ICWSM 2011, Barcelona, Catalonia, Spain, July 17-21, 2011*. The AAAI Press, pages 554–557.
- Ehsan Mohammady Ardehaly and Aron Culotta. 2015. Inferring latent attributes of Twitter users with label regularization. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Denver, Colorado, pages 185–195.
- Shirui Pan, Jia Wu, Xingquan Zhu, Chengqi Zhang, and Yang Wang. 2016. Tri-party deep network representation. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*. pages 1895–1901.
- Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, KDD '14, pages 701–710.
- Jordan B. Pollack. 1990. Recursive distributed representations. *Artificial Intelligence* 46(1-2):77–105.
- Daniel Preotiuc-Pietro, Vasileios Lamos, and Nikolaos Aletras. 2015. An analysis of the user occupational class through Twitter content. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Beijing, China, pages 1754–1764.
- Qiao Qian, Bo Tian, Minlie Huang, Yang Liu, Xuan Zhu, and Xiaoyan Zhu. 2015. Learning tag embeddings and tag-specific composition functions in recursive neural network. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Beijing, China, pages 1365–1374.
- Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar, and Hansen Andrew Schwartz. 2014. Developing age and gender predictive lexica over social media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, pages 1146–1151.

- Richard Socher, Eric H. Huang, Jeffrey Pennington, Andrew Y. Ng, and Christopher D. Manning. 2011. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*. Curran Associates Inc., USA, NIPS' 11, pages 801–809.
- Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, WWW ' 15, pages 1067–1077.
- Tomoki Taniguchi, Shigeyuki Sakaki, Ryosuke Shigenaka, Yukihiro Tsuboshita, and Tomoko Ohkuma. 2015. A weighted combination of text and image classifiers for user gender inference. In *Proceedings of the Fourth Workshop on Vision and Language*. Association for Computational Linguistics, Lisbon, Portugal, pages 87–93.
- Jose Manuel Delgado Valdes, Jacob Eisenstein, and Munmun De Choudhury. 2015. Psychological effects of urban crime gleaned from social media. In *Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*. AAAI Press, Menlo Park, California, pages 598–601.
- Svitlana Volkova. 2014. Twitter data collection: Crawling users, neighbors and their communication for personal attribute prediction in social media. *Center for Language and Speech Processing, Johns Hopkins University*.
- Svitlana Volkova, Glen Coppersmith, and Benjamin Van Durme. 2014. Inferring user political preferences from streaming communications. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, Maryland, pages 186–196.
- Stephen Wan and Cécile Paris. 2014. Improving government services with social media feedback. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*. ACM, New York, NY, USA, IUI ' 14, pages 27–36.
- Fei Wang and Changshui Zhang. 2006. Label propagation through linear neighborhoods. In *Proceedings of the 23rd International Conference on Machine Learning*. ACM, New York, NY, USA, ICML '06, pages 985–992.
- Qiongkai Xu, Qing Wang, Chenchen Xu, and Lizhen Qu. 2017. Collective vertex classification using recursive neural network. *arXiv preprint arXiv:1701.06751*.
- Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y. Chang. 2015. Network representation learning with rich text information. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*. AAAI Press, IJCAI' 15, pages 2111–2117.
- Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. 2016. Revisiting semi-supervised learning with graph embeddings. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*. pages 40–48.
- Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one* 11(3):e0150989.