Using Relational Framing to Characterize Geographical Differences in Views on Immigration

Edward Newell and **Derek Ruths**
School of Computer Science
McGill University

Abstract

We introduce a new method for text analysis, and demonstrate its use in characterizing geographical differences in views on immigration in the United States and relations with Latin America. Our approach adds to a growing literature on modeling bias, positioning, and framing in text.

A key innovation is the modeling of relationships explicitly depicted between entities in the text. Starting from a text source, we generate an *entity-network*, in which entities are nodes and relationships are edges. We find the method enables deep analysis of social media streams through its ability to generate highly structured output with links to a knowledge base. This enables flexible, targeted queries that can consider relationships among both concrete entities and high-level abstractions depicted in the text.

When our method is applied to content from geo-located Twitter users, we find that US residents near the border of Mexico show a greater concern for concrete issues affecting immigration in the present, and the involvement of currently-serving politicians. In contrast, residents far from the border have a higher tendency to engage these issues in connection to election-time campaigns. Residents near the border more readily view interactions between the US and Latin America as a discussion between politicians on both sides, while those further from the border are less likely to recognize Latin American politicians as actors.

Our findings agree with the idea that the US population near the border with Mexico is both more directly impacted by immigration from Latin America, and has a higher fraction of people of Latin American descent, who will tend to identify more with Latin American culture. This provides support for our approach as a technique for population studies through large-scale text analysis.

1 Introduction

In social media, users express positions on a variety of topics. Characterizing such views offers an exciting and valuable opportunity to better understand populations of users as well as the larger offline populations for which they may proxy.

Past work on modeling bias, positions, and framing in text has focused largely on discovering the entities and concepts that characterize a particular point-of-view (e.g., [2, 9]). Such approaches, however, do not consider the relational structure of a text. In this paper, we make the case that this relational structure (which we call *relational framing*) is important to a complete characterization of the viewpoint expressed in a text. We show this importance through an analysis of how Twitter users' positions on US-Latin American immigration vary as a function of their distance from the US-Mexico border.

Formally, we model the relational framing of a text as the network of entities mentioned in the text. Edges are the relationships explicitly asserted by the text among the entities. This structure is learned using a natural language processing pipeline which involves a number of stages including entity recognition, entity linking, co-reference resolution, and interaction inference.

We demonstrate the utility of our approach to modeling relational framing by showing how it can be used to learn how views on the issue of US-Latin American immigration vary geographically. In recent months, this topic has become a major talking-point among US presidential hopefuls. Moreover, the topic has been studied in depth through other means, making our demonstration both a way of validating our approach against existing work and showing how, more generally, it can be applied.

As a technique for demographic research, we hope relational framing and our approach to modeling it will enable researchers to ask (and answer) nuanced questions about the views and ideologies of online populations. Such questions might include processes of social integration of immigrants, attitudes toward marriage and reproduction, or the views of citizens toward proposed or newly-implemented policies. Combined with established methods for inferring latent demographic variables of online users such as age, gender, and location — our proposed approach will allow demographers a new lens through which to study how demography and discourse shape one another.

In the next section, we provide a broad overview of the approach. We then apply it to explore geographical differences in views of US residents and examine the results.

2 Overview of Methodology

The Twitter API serves as the primary data source in this work. The data collection and processing can be described broadly as follows. In the first stage (Figure 1A), we identify a set of Twitter users based on immigration-related keyword and hashtag searches. We geolocate these users, which, given a focus on US residents, allows us to resolve users' locations to specfic US states. In the second phase (Figure 1B), we obtain each geolocated user's most recent 400 tweets, and then extract all the links (URLs) found in those tweets. We follow the links, and extract the textual content found on the target web pages. This content then becomes the input to our novel analysis approach, which applies a series of natural language processing techniques to identify entities

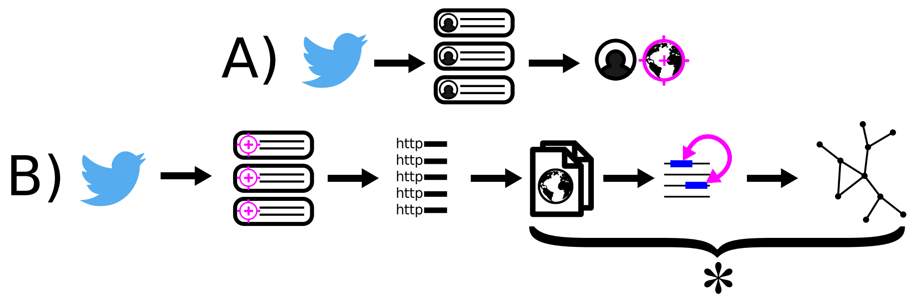

Figure 1: High-level steps of data collection and processing. In the first phase **(A)** data collection begins by collecting tweets using immigration-related keywords, followed by geolocation of the users. In the second phase **(B)** the most recent 400 tweets from geolocated users, and the links found in the tweets are extracted. Text content from the target websites is processed using natural language processing and information extraction to extract entities and interactions and generate entity-interaction network. The combination of steps highlighted by the asterisk represents the core novelty in the approach.

and interactions in the text, and then uses the results to assemble a network representation. Based on the location of the user who linked a given web-page, we create distinct network representations associated to specific geographical regions.

The novelty of our approach lies in the final steps in which the fetched text content is converted into networks of entity interactions (marked by the asterisk, Figure 1). Although depicted pictorially as a small part of the data collection and processing, these steps are far more computationally intensive and combine several state of the art natural language processing and information extraction tools. These steps can be applied to any raw text source, and are not limited to the analysis of Twitter or content scraped from the web.

Our choice to use the content found in web pages, rather than the content of tweets themselves, is motivated by the fact that web pages contain longer passages of coherent content. Especially in the limited space afforded by tweets, we find that users frequently express their views by sharing links to articles and other content. We take this linked content as a proxy for users' views based on the fact that the user has explicitly flagged it as noteworthy and worth sharing.

At the core of our approach is a design choice which establishes interactions between entities as a basic unit of analysis. In the kinds of documents that we expect to focus on (typically news articles and blog posts) some coherent topic is usually discussed, often involving the depiction of a series of events. This generally involves the mention of named entities—specific people, places, or organizations that can be positively identified by their proper names—which are depicted as interacting and influencing one another. In the event extraction literature, events are typified by the participants involved, together with a time

immigrants	9999	#immigrationreform	1291	#immigrants	199
immigration	9996	#immigration	1098	migrant workers	196
immigrant	4793	#illegalaliens	983	#migrantworkers	99
illegal aliens	3195	#illegalalien	631	#immigrant	98
illegal alien	2997	foreign worker	396	#deport	72
border security	2891	foreign workers	395	#stoph1b	40
h1b	2691	#bordersecurity	372	#noh1b	34
deport	1499	#borderpatrol	359	#foreignworkers	14
#h1b	1487	migrant worker	295	#migrantworker	6
border patrol	1399	#illegalimmigration	285	#foreignworker	1

Table 1: Keywords and hashtags used to find Twitter users who discuss immigration issues pertaining to the US.

and a location [6, 21]. Entities (people, places, and organizations) represent the active agents in a text, and their interactions provide a signature of the events that frame the story presented in the text—what we call the text's *relational framing*. Entities and interactions are also highly specific: entities can be cross-referenced with a knowledge base, and the combinations of entities appearing in interactions is highly indicative not only of the topic being discussed, but also the specific focus, and how the entities are attributed roles and associated with one another.

The extraction of entities and interactions begins with the application of the CoreNLP tool suite [16], which provides tokenization, dependency parsing, named entity recognition, and coreference resolution. We then employ the AIDA system [11] to link named entities to the YAGO [10], a knowledge base built from Wikipedia[1], WordNet [17], and GeoNames[2]. Linkage to YAGO provides a common point of reference to associate mentions of the same entity in separate documents, but also greatly enriches the entity graph metadata and facts about the entities. This latter benefit makes it possible to query specific entities based on the properties like place of birth, profession, or membership on a team. Finally, we use the ReVerb [8] and PATTY [18] systems to uncover the instances where interactions are depicted between entities. The output from this analysis is a network with entities as nodes and interactions as edges.

For the current analysis, the scope of raw data to be analyzed is determined by the keyword and hashtag searches that we use to select Twitter user accounts. We began by simply searching for the keyword "immigration". This returned tweets on many topics, including issues related to immigration in the US and the refugee crisis due to events centered on Syria and Iraq. Reading these tweets, we identified 30 keywords that preferentially returned tweets about immigration in the US (See Table 1). We collected all users accounts returned in searches with these keywords and hashtags.

The task of geolocation has been studied extensively elsewhere [1, 4, 5, 7, 12–

[1] www.wikipedia.org
[2] www.geonames.org

15, 19, 20]. For this step, we simply rely on the Macromeasures[3] service. This services uses techniques described in the literature in addition to proprietary techniques.

Once all of the text from web pages linked by users has been processed, we obtain entity-interaction networks associated to particular geographical regions. For the purpose of our demonstration, we separated US residents into two groups: those from states that border Mexico, and those from states far from Mexico, which, henceforth, we will refer to as *border residents* and *distant residents*. Among distant residents we include residents of all states except (a) states that border Mexico, and (b) states that border a state bordering Mexico (e.g. Nevada is excluded). This provides geographically distinct groups which we expect to differ with regard to their views toward immigration and relations between the US and Latin America.

3 Results

Based on the process described above, we obtained two entity networks, one serving as a proxy for views held by residents of the states bordering Mexico (*border residents*), and the other for views held by residents of states far from Mexico (*distant residents*). These networks have on the order of $10,000$ entities and tens of thousands of relationships (see Table 2).

The first observation we make is that, since the distant residents represent a larger population, our sample contains more users in that group, and hence the network for distant residents has more nodes and edges. Before any meaningful comparisons can be made, the networks must be normalized. Among ways to approach normalization, we can: (1) divide the weight of each edge (which encodes the strength of relationship between the entities) by the total weight of all edges in the network, which we will call *fractional normalization*, and (2) randomly subsample the textual mentions of interactions used to construct the larger network, thereby constructing both networks from the same number of input interaction mentions, which we will call *subsampling normalization*. Fractional normalization has an advantage in that it does not discard information. However, fractional normalization does not adjust the number of nodes and edges, so isn't suitable for topological comparisons; for that purpose, subsampling is appropriate.

Although we selected users based on their use of immigration-related keywords, their tweets and links will cover many topics. We can select interactions of interest by making use of the YAGO knowledge base. For example, in YAGO, Barack Obama is attributed various classifications, including `People_from_Honolulu,_Hawaii`, `African-American_non-fiction_writers`, and `Presidents_of_the_United_States`. For our first comparison, we select all interactions that occur between a US-affiliated entity and a LA (Latin America)-affiliated entity. (For brevity, we will henceforth refer to Latin America as LA.)

[3]macromeasures.com

Subnetwork (normalization)	Statistic	Border residents	Distant residents
Full (none)	Entities	9509	16259
	Unique relationships	21976	41993
	Total relationship weight (count)	145106	372632
US-LA (fractional)	Entities	671	1049
	Unique relationships	867	1478
	Total relationship weight (ppm)	27197.2	25643.0
US-LA (subsampled)	Entities	671	649
	Unique relationships	867	817
	Total relationship weight (count)	3934	3696

Table 2: Statistics for entity networks obtained for border residents and distant residents. The full networks and subnetworks restricted to relationships between US- and LA-associated entities are shown. Total edge weight for the fractionally normalized networks is expressed as a fraction of the corresponding full networks' total weight in parts per million notation.

This selection of relationships defines subnetworks within the original networks. The properties of these subnetworks (which were created after subsampling or fractional normalization on the full networks) provide a view of the emphasis accorded to these relationships by border residents and distant residents (see Table 2). For both normalization methods, we find that the total weight of relationships—and hence the total emphasis placed on US-LA interactions—is larger for border residents than for distant residents. After normalization by subsampling, the US-LA subnetwork of border residents is also larger in terms of number of entities and relatinoships. Overall, border residents direct more attention to relationships between US-affiliated and LA-affiliated entities.

We can look closer at where the two populations direct their attention, by looking at particular entity interactions. Still taking a relatively coarse view, we can look at the interactions with the US on side and Mexico or Central American countries on the other (see Table 3). Here we are refering to interactions depicted between the countries themselves, not between all entities affiliated with them.

The weights of these relationships illustrate a few notable trends. On balance, border residents attribute more weight to these relationships than do distant residents. However, the distributions of weights are quite varied. The relationships between El Salvador and Guatemala, are attributed more than 5 and 8 times more weight by border residents, whereas Panama receives 18 ppm from distant residents but is never explicitly mentioned in a relationship with the United States by border residents.

Furthermore, from the country-level interactions, we do not see all of the interactions between other affiliated entitites that might proxy for the countries in the source texts. Taking the case of Panama, if we consider relationships between all Panama-affiliated and US-affilated entities, we find that border res-

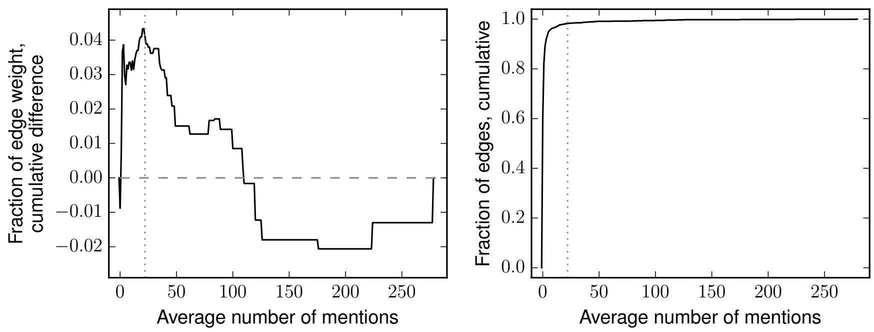

Figure 2: The left plot shows how edge weight is distributed in the US-LA plots (under subsampling normalization) for border and distant residents. The value plotted shows the difference in the cumulative weight accounted for by edges whose individual weight is less than or equal to the value on the horizontal axis. Most of the weight for the border residents' network is found in edges having fewer than 23 mentions (vertical line) mentions, whereas distant residents have more weight on stronger edges. The right plot shows the cumulative fraction of edges from both graphs: 98% have a weight of less than 23 mentions.

idence do attribute weight to these relationships, in fact, more than do distant residents (55.3 ppm compared to 37.6 ppm).

Relationships can arise at various levels, and the relationships between certain entities may be framed in terms of relationships between other, affiliated entities. This can take place in the form of metonymy, such as when "Washington" is used to refer to the US government. But this can also arise when entities can be viewed as hierarchically related. For many purposes, it may make sense to think of a country as made up of all of its citizens, politicians, organizations, geographical features, etc. Considering the relationships the country engages in may involve considering the relationships that all these "contained" entities engage in. (It was possible to consider all of these relationships in forming the US-LA networks in Table 2 using YAGO along with gazetteers.)

When describing events, there is a certain amount of freedom of choice in deciding which interactions and associations to depict, as well as in deciding which entities represent a given group or camp. These choices carry information about how the events are framed. A social media user who is very interested in a given topic is more likely to link in-depth articles covering that topic. These in turn are more likely to describe detailed and nuanced interactions which may be less widely known. In contrast, more perfunctory coverage of the same topic will likely relate only the most prominent entities and interactions. Since border residents have a larger fraction of people of Latin American descent [3], we should expect to see extra weight attributed to the less prominent interactions involving LA entities.

A direct comparison of where the weight is distributed in the US-LA net-

Subnet	**Border residents**	**Distant residents**
Mexico	2157.0	1714.2
Cuba	705.1	810.0
Honduras	145.2	96.9
Nicaragua	13.8	18.8
Panama	0.	18.8
Guatemala	89.8	16.2
El Salvador	117.5	13.4
Costa Rica	13.8	24.2
Belize	0.	2.6
Total	3249.1	2852.5

Table 3: Strength of relationships between the indicated country, and the United States, as found in the US-LA fractionally normalized subnetworks for border and distant residents. Edge weights are in units of parts per million.

works for border and distant residents can be made using a suitably prepared plot such as the one in Figure 2, left. The plot is based on the cumulative fraction of weight, accounted for by edges up to a given individual weight, for each network. The difference between these quantities is what is plotted; thus the curve shows how much more of the weight in the border residents' network is accounted for by edges up to a given weight, compared to in the distant residents' network.

Initially the curve rises almost monotonically, indicating that more weight in the border residents' network is allocated to less prominent edges. For relationships having a weight greater than 23 mentions, more weight is found in the distant residents' network. One thing which is not immediately apparent is that, as shown in the Figure 2 right, more than 98% of edges have a weight of 23 or fewer mentions. In other words, border residents' allocate more attention to less prominent interactions, while distant residents focus the balance of this weight on the 2% most prominent interactions.

Having shown how attention varies with prominence, we compare residents' focus on different categories of entities. To do so, we make use of entity attributes from the YAGO knowledge base. We have already seen that when we compared the subnetworks consisting of interactions between US- and LA-affiliated entities, we see that border residents attribute more weight to these interactions than distant residents (about 6% more). If we further restrict our focus to politicians the discrepancy increases with border residents attributing 75% more weight to interactions between US and LA politicians, (see Table 4). This shows that the differences in how residents view the relationships between LA and US are particularly pronounced in the political dimension.

Interestingly, the excess emphasis that border residents place on the political dimension of US-LA relations seems to be particularly oriented to LA-

US entities	**LA entities**	**Border b residents**	**Distant d residents**	$(b-d)/d$ (%)
All	All	27,197.2	25,643.0	6.0
Politicians	All	5,136.6	4,760.5	7.9
All	politicians	1,265.1	1,079.1	17.2
politicians	politicians	207.4	118.4	75.2
Presidential candidates	All	5247.2	5309.4	-1.2
Barack Obama	All	1382.7	1324.0	4.4
Key depts. and services	All	504.7	196.4	156.9

Table 4: Fraction of total edge weight (expressed in parts per million), in the networks derived from border residents and distant residents, found in interactions between US- and LA-affiliated entities of the indicated type.

`United_States_Department_of_Justice`	`United_States_Customs_Service`
`Washington_Office_on_Latin_America`	`Cuban_Refugee_Adjustment_Act`
`Drug_Enforcement_Administration`	`United_States_Census_Bureau`
`Americans_for_Legal_Immigration`	`United_States_Border_Patrol`
`National_Border_Patrol_Council`	

Table 5: US departments, Acts, and organizations related to border security, and immigration legislation

politicians. When we look at subnetworks based on relationships between US politicians and any LA-affiliated entity, or the subnetworks based on LA politicians and any US-affiliated entity, the greater difference is seen when restricting focus to LA politicians. One possible explanation is that residents near the border more readily view relations between LA and the US in terms of a conversation between politicians, with distant residents preferentially framing the interaction around US political actors.

We can subdivide the political entities further to gain more insight. This study was conducted at a time when the presidential candidates were campaigning for nomination by the Democratic and Republican parties in the United states, and so much of the interactions depicted in the networks may arise out of election-time campaigning. Looking at only the interactions that the incumbent, Barack Obama, engages in with LA entities, we still observe greater emphasis by border residents (Table 4). But, when we focus on interactions between presidential candidates and LA entities, we find the reverse: distant residents emphasize the candidates' relations with LA entities more. This surprising given border residents generally emphasized politicians' relations among US-LA interactions more. But this reversal in emphasis is consistent with a situation in which border residents focus more on concrete, ongoing issues, while distant residents consider US-LA interactions due to their incorporation in campaign platforms.

Subnet	Statistic	Border residents	Distant residents
LA-LA fractional	Entities	671	1049
	Unique relationships	867	1478
	Total edge weight (ppm)	4016.7	3027.4
LA-LA subsampled	Entities	178	173
	Unique relationships	213	208
	Total edge weight	581	428
	Clustering (%)	4.4	7.6

Table 6: Statistics for entity subnetworks obtained by considering only interactions between LA-affiliated entities, normalized using fractional and subsampled normalization. The total edge weight for the fractionally normalized network is expressed as a fraction of the total network's weight in parts per million notation. Global clustering coefficient is calculated for the networks normalized by subsampling.

We can focus more specifically on issues pertaining to immigration by inspecting the relationships between LA entities and relevant US departments, acts, and organizations whose purposes relate to immigration (listed in Table 5). These entities have a material affect on border residents, so their relationships should get much stronger emphasis from border residents than distant residents. This is indeed what we find: relationships to these entities receive about 2.5 times as much emphasis by border residents.

So far, we have discussed dyadic relations, in which we consider one entity associated to Latin America and one entity associated to the US. This lets us ask conceptually straightforward questions about the output of our analysis. However, the models of entities and relationships that we obtain can capture local and global structure beyond dyadic relations.

We now consider the subnetworks formed by relationships among Latin American entities. There are good reasons to expect the LA subgraph to differ for our two populations. Since border residents contain a higher fraction of people of Latin American descent [3] who are more likely to identify with Latin America, we expect to find a fuller subnetwork of relationships among Latin American entities. Indeed, the total weight in border residents' LA-LA network is greater, and, when subsampling based normalization is used, there are more nodes and edges in the border residents' network (see Table 6).

We also measured the clustering coefficient, but it has a more subtle interpretation. While we might expect a denser network with a higher clustering coefficient, we have already seen evidence that greater attention to a given issue can lead to the inclusion of *less* prominent relationships with regard to that issue. One would in turn expect this to drive *down* the clustering coefficient. In the case of the subsampling-based normalized LA-LA networks, we do observe substantially smaller clustering coefficient for the border residents' network.

4 Conclusion

We have demonstrated a new text analysis technique that extracts the relationships mentioned between entities. As we have shown, the relational structure we derive makes it possible to investigate how a body of text frames particular topics in terms of the interactions between agents in the world.

We have demonstrated the approach through a large-scale analysis of how Twitter users' views toward immigration vary geographically. Our analysis reveals broad trends, as well as highly resolved details about these views. The analysis was possible because the networked representation we derive is detailed yet amenable to making high-level abstract queries. These characteristics make the approach well-suited to the large-scale studies of populations using social media.

References

[1] Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web*, pages 61–70. ACM, 2010.

[2] David Bamman and Noah A Smith. Open extraction of fine-grained political statements. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015.

[3] U.S. Census Bureau. American FactFinder, 2014. *Accessed 5 March 2016.*

[4] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 759–768. ACM, 2010.

[5] Ryan Compton, David Jurgens, and David Allen. Geotagging one hundred million twitter accounts with total variation minimization. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 393–401. IEEE, 2014.

[6] Agata Cybulska and Piek Vossen. Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In *LREC*, pages 4545–4552, 2014.

[7] Clodoveu A Davis Jr, Gisele L Pappa, Diogo Rennó Rocha de Oliveira, and Filipe de L Arcanjo. Inferring the location of twitter messages based on user relationships. *Transactions in GIS*, 15(6):735–751, 2011.

[8] Anthony Fader, Stephen Soderland, and Oren Etzioni. Identifying relations for open information extraction. In *Proceedings of the Conference of Empirical Methods in Natural Language Processing (EMNLP '11)*, Edinburgh, Scotland, UK, July 27-31 2011.

[9] Swapna Gottipati, Minghui Qiu, Yanchuan Sim, Jing Jiang, Noah A Smith, and A Model. Learning topics and positions from debatepedia supplementary material. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013.

[10] Johannes Hoffart, Fabian M Suchanek, Klaus Berberich, and Gerhard Weikum. Yago2: A spatially and temporally enhanced knowledge base from wikipedia. *Artificial Intelligence*, 194:28–61, 2013.

[11] Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. Robust disambiguation of named entities in text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 782–792. Association for Computational Linguistics, 2011.

[12] David Jurgens. That's what friends are for: Inferring location in online social media platforms based on social relationships. *ICWSM*, 13:273–282, 2013.

[13] David Jurgens, Tyler Finethy, James McCorriston, Yi Tian Xu, and Derek Ruths. Geolocation prediction in twitter using social networks: a critical analysis and review of current practice. In *Proceedings of the 9th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2015.

[14] Longbo Kong, Zhi Liu, and Yan Huang. Spot: Locating social media users based on social network context. *Proceedings of the VLDB Endowment*, 7(13):1681–1684, 2014.

[15] Rui Li, Shengjie Wang, Hongbo Deng, Rui Wang, and Kevin Chen-Chuan Chang. Towards social user profiling: unified and discriminative influence model for inferring home locations. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1023–1031. ACM, 2012.

[16] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60, 2014.

[17] George Miller and Christiane Fellbaum. Wordnet: An electronic lexical database, 1998.

[18] Ndapandula Nakashole, Gerhard Weikum, and Fabian Suchanek. Patty: a taxonomy of relational patterns with semantic types. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1135–1145. Association for Computational Linguistics, 2012.

[19] Reid Priedhorsky, Aron Culotta, and Sara Y Del Valle. Inferring the origin locations of tweets with quantitative confidence. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 1523–1536. ACM, 2014.

[20] Dominic Rout, Kalina Bontcheva, Daniel Preoţiuc-Pietro, and Trevor Cohn. Where's@ wally?: a classification approach to geolocating users based on their social ties. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 11–20. ACM, 2013.

[21] Congle Zhang, Stephen Soderland, and Daniel S Weld. Exploiting parallel news streams for unsupervised event extraction. *Transactions of the Association for Computational Linguistics*, 3:117–129, 2015.