

Annotation Scheme for Social Network Extraction from Text

Apoorv Agarwal

Computer Science Department
Columbia University
New York, U.S.A.

apoorv@cs.columbia.edu

Owen Rambow

CCLS
Columbia University
New York, U.S.A.

rambow@ccls.columbia.edu

Rebecca J. Passonneau

CCLS
Columbia University
New York, U.S.A.

becky@cs.columbia.edu

Abstract

We are interested in extracting social networks from text. We present a novel annotation scheme for a new type of event, called *social event*, in which two people participate such that at least one of them is cognizant of the other. We compare our scheme in detail to the ACE scheme. We perform a detailed analysis of inter-annotator agreement, which shows that our annotations are reliable.

1 Introduction

Our task is to extract a social network from written text. The extracted social network can be used for various applications such as summarization, question answering, or the detection of main characters in a story. We take a “social network” to be a network consisting of individual human beings and groups of human beings who are connected to each other through various relationships by the virtue of participating in events. A text can describe a social network in two ways: explicitly, by stating the type of relationship between two individuals (Example ??); or implicitly, by describing an event which creates or perpetuates a social relationship (Example 2). We are interested in the implicit description of social relations through events. We will call these types of events *social events*. Crucially, many social relations are described in text largely implicitly, or even entirely implicitly. This paper presents an annotation project for precisely such social events.

To introduce the terminology and conventions we use throughout the paper, consider the following Example 2. In this example, there are two entities: **Iraqi officials** and **Timothy McVeigh**. These entities are present in text as nominal and named entity mentions respectively (within [...]). Furthermore, these entities are related by

an event, whose type we call INR.NONVERBAL-NEAR (a non-verbal interaction that occurs in physical proximity), and whose textual mention is the extent (or span of text) *provided money and training*.¹

- (1) [[Sharif]’s {wife} Tahari Shad Tabussum], 27, (...) made no application for bail at the court, according to local reports. **PER-SOC**
- (2) The suit claims [Iraqi officials] {provided money and training} to [convicted bomber Timothy McVeigh] (...). **INR.Nonverbal-Near**

One question that immediately comes to mind is how would these annotations be useful? Let us consider the problem of finding the hierarchy of people in the Enron Email corpus (Klimt and Yang, 2004; Diesner et al., 2005). Much work to solve this problem has focused on using social network analysis algorithms for calculating the graph theoretical quantities (like degree centrality, clustering coefficient (Wasserman and Faust, 1994)) of people in the email sender-receiver network (Rowe et al., 2007). Attempts have been made to incorporate the content of emails usually by using topic modeling techniques (McCallum et al., 2007; Pathak et al., 2008). These techniques consider a distribution of words in emails to classify the interaction between people into topics and then cluster together people that talk about the same topic. Researchers also map relationships among individuals based on their patterns of word use in emails (Keila and Skillicorn, 2005). But these techniques do not attempt to create an accurate social network in terms of interaction or cognitive states of people. In comparison, our data allows

¹Throughout this paper we will follow this representation scheme for examples – entity mentions will be enclosed in square brackets [...] and relation mentions will be enclosed in set brackets {...}

Sender → Receiver	Email content
Kate → Sam	[Jacob], the City attorney had a couple of questions which [I] will {attempt to relay} without having a copy of the documents.
Sam → Kate, Mary	Can you obtain the name of Glendale’s bond counsel (lawyer’s name, phone number, email, etc.)?
Kate → Sam	Glendale’s City Attorney is Jacob. Please let [me] {know} if [you] need anything else.
Mary → Sam	I do not see a copy of an opinion in the file nor have we received one since [I] {sent} the execution copies of the ISDA to [Jacob].
Kate → Jacob	Jacob, could you provide the name, phone number, etc. of your bond council for our attorney, Sam?
Kate → Sam	[I] will {work on this for} [you] - and will be in touch.

Figure 1: An email thread from the Enron Email Corpus. (For space concerns some part of the conversation is removed. The missing conversation does not affect our discussion.)

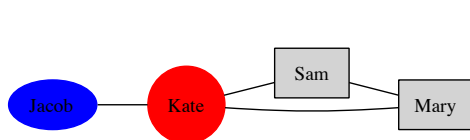


Figure 2: Network formed by considering email exchanges as links. Identical color or shape implies structural equivalence. Only **Sam** and **Mary** are structurally equivalent

for such a technique to be created. This is because our annotations capture interactions described in the content of the email such as face-to-face meetings, physical co-presence and cognizance.

To explore if this is useful, we analyzed an Enron thread which is presented in Figure 1. Figure 2 shows the network formed when only the email exchange is considered. It is easy to see that **Sam** and **Mary** are structurally equivalent and thus have the same role and position in the social network. When we analyze the content of the thread, a link gets added between **Mary** and **Jacob** since **Mary** in her email to **Sam** talks about sending something to **Jacob**. This link changes the roles and positions of people in the network. In the new network, Figure 3, **Kate** and **Mary** appear structurally equivalent to each other, as do **Sam** and **Jacob**. Furthermore, **Mary** now emerges as a more important player than the email exchange on its own suggests. This rather simple example is an indication of the degree to which a link may affect the social network analysis results. In emails where usually a limited number of people are involved, getting an accurate network seems to be crucial to the hierarchal analysis.

There has been much work in the past on an-

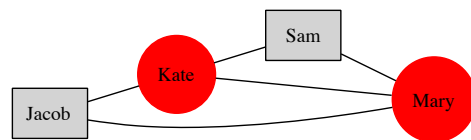


Figure 3: Network formed by augmenting the email exchange network above with links that occur in the content of the emails. Now, **Kate** and **Mary** are structurally equivalent, as are **Sam** and **Jacob**.

notating entities, relations and events in free text, most notably the ACE effort (Doddington et al., 2004). We intend to leverage this work as much as possible. The task of social network extraction can be broadly divided into 3 tasks: 1) entity extraction; 2) social relation extraction; 3) social event extraction. We are only interested in the third task, social event extraction. For the first two tasks, we can simply use the annotation guidelines developed by the ACE effort. Our social events, however, do not clearly map to the ACE events: we introduce a comprehensive set of social events which are very different from the event annotation that already exists for ACE. This paper is about the annotation of social events.

The structure of the paper is as follows. In Section 2 we present a list of social relations that we annotate. We also talk about some design decisions and explain why we took them. We compare this annotation to existing annotation, notably the ACE annotation, in Section 3. In section 4 we present the procedure of annotation. Section 5 gives details of our inter-annotator agreement calculation procedure and shows the inter-annotator agreement on our task. We conclude in section 6

and mention future direction of research.

2 Social Event Annotation

In this section we define the social events that the annotators were asked to annotate. Here, we are interested in the meaning of the annotation; details of the annotation procedure can be found in Section 4. Note that in this annotation effort, we do not consider issues related to the truth of the claims made in the text we are analyzing — we are interested in finding social events whether they are claimed as being true, presented as speculation, or presented as wishful thinking. We assume that other modules will be able to determine the factive status of the described social events, and that social events do not differ from other types of events in this respect.

A social event is an event in which two or more entities relate, communicate or are associated such that for at least one participant, the interaction is **deliberate** and **conscious**. Put differently, at least one participant must be aware of relating to the other participant. In this definition, what constitutes a social relation is an aspect of cognitive state: an agent is aware of being in a particular relation to another agent. While two people passing each other on a street without seeing each other may be a nice plot device in a novel, it is not a social event in our sense, since it does not entail a social relation.

Following are the four types of social events that were annotated:²

Interaction event (INR): When both entities participating in an event have each other in their cognitive state (i.e., are aware of the social relation) we say they have an INR relation. The requirement is actually deeper: it extends to the transitive closure under mutual awareness, what in the case of belief is called “mutual belief”. An INR event could either be of sub-type VERBAL or NONVERBAL. Note that a verbal interaction event does not mean that all participants must actively communicate verbally, it is enough if one participant communicates verbally and the others are aware of this communication.³ Furthermore, the interaction can be in physical proximity or from a distance. Therefore, we have further subtypes of

²Details of the annotation guidelines can be found in the unpublished annotation manual, which we will refer to in the final version of the paper.

³For this reason we explicitly annotate legal events as VERBAL because legal interactions usually involve words

INR relation: NEAR and FAR. In all, INR has four subtypes: VERBAL-NEAR, VERBAL-FAR, NONVERBAL-NEAR, NONVERBAL-FAR. Consider the following Example (3). In this sentence, our annotators recorded an INR.VERBAL-FAR between entities **Toujan Faisal** and the **committee**.

- (3) [Toujan Faisal], 54, {said} [she] was {informed} of the refusal by an [Interior Ministry committee] overseeing election preparations. **INR.Verbal-Far**

As is intuitive, if one person *informs* the other about something, both have to be cognizant of each other and of the *informing* event. Also, the event of *informing* involves words, therefore, it is a verbal interaction. From the context it is not clear if **Toujan** was informed personally, in which case it would be a NEAR relation, or not. We decided to default to FAR in case the physical proximity is unclear from the context. We decided this because, on observation, we found that if the author of the news article was reporting an event that occurred in close proximity, the author would explicitly say so or give an indication. INR is the only relation which is bi-directional.

Cognition event (COG): When only one person (out of the two people that are participating in an event) has the other in his or her cognitive state, we say there exists a cognition relationship between entities. Consider the aforementioned Example (3). In this sentence, the event *said* marks a COG relation between **Toujan Faisal** and the **committee**. This is because, when one person talks about the other person, the other person must be present in the first person’s cognitive state. COG is a directed event from the entity which has the other entity in its cognitive state to the other entity. In the example under consideration, it would be from **Toujan Faisal** to the **committee**. There are no subtypes of this relation.

Physical Proximity event (PPR): We record a PPR event when both the following conditions hold: 1) exactly one entity has the other entity in their cognitive state (this is the same requirement as that for COG) and 2) both the entities are physically proximate. Consider the following Example (4). Here, one can reasonably assume that **Asif Muhammad Hanif** was aware of being in physical proximity to the **three people** killed, while the inverse was not necessarily true.

- (4) [Three people] were killed when (...), [Asif Muhammad Hanif], (...), {detonated explosives strapped to [his] body} **PPR**

PPR is a directed event like COG. There are no subtypes of this relation. Note that if there exists a PPR event then of course there would also be a COG event. In such cases, the PPR event subsumes COG, and we do not separately record a COG event.

Perception event (PCR): The Perception Relationship is the distant equivalent of the Physical Proximity event. The point is not physical distance; rather, the important ingredient is the awareness required for PPR, except that physical proximity is not required, and in fact physical distance is required. This kind of relationship usually exists if one entity is watching the other entity on TV broadcast, listening to him or her on the radio or using a listening device, or reading about the other entity in a newspaper or magazine etc. Consider the following Example (5). In this example, we record a PCR relation between the **pair** and the **Nepalese babies**. This is because, the babies are of course not aware of the pair. Moreover, the pair heard about the babies so there is no physical proximity. It is not COG because there was an explicit external information source which brought the babies to the attention of the pair.

- (5) [The pair] flew to Singapore last year after {hearing} of the successful surgery on [Nepalese babies] [Ganga] and [Jamuna Shrestha], (...). **PCR**

PCR is a directed event like COG. There are no subtypes of this relation. Note that if there exists a PCR event then we do not separately record a COG event.

Figure 4 represents the series of decisions that an annotator is required to take before reaching a terminal node (or an event annotation label). The interior nodes of the tree represent questions that annotators answer to progress downwards in the tree. Each question has a binary answer. For example, the first question the annotators answer to get to the type and subtype of an event is: “Is the relation directed (1-way) or bi-directional (2-way)?” Depending on the answer, they move to the left or the right in the tree respectively. If its a 2-way relation, then it has to one of the sub-types of INR because only INR requires that both entities be aware of each other.

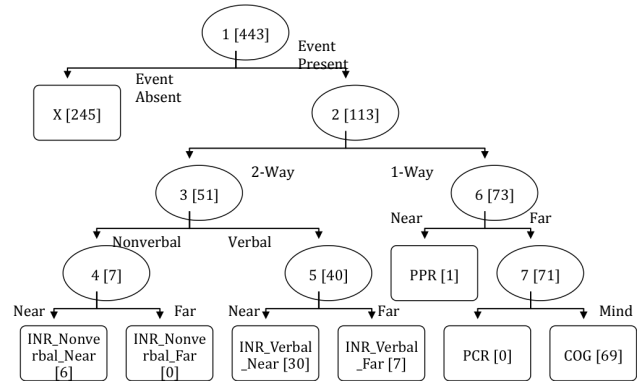


Figure 4: Tree representation of decision points for selecting an event type/subtype out of the list of social events. Each decision point is numbered for easy reference. We refer to these numbers later when we present our results. The numbers in braces ([...]) are the number of examples that reach a decision point.

3 Comparison Between Social Events and ACE Annotations

In this section, we compare our annotations with existing annotation efforts. To the best of our knowledge, no annotation effort has been geared towards extracting social events, or towards extracting expressions that convey social relations in text. The Automated Content Extraction (ACE) annotations are the most similar to ours because ACE also annotates Person Entities (PER.Individual, PER.Group), Relations between people (PER-SOC), and various types of Events. Our annotation scheme is different, however, because the focus of our event annotation is on events that occur only between people. Furthermore, we annotate text that expresses the cognitive states of the people involved, or allows the annotator to infer it. Therefore, at the top level of classification we differentiate between events in which only one entity is cognizant of the other versus events when both entities are cognizant of each other. This distinction is, we believe, novel in event or relation annotation. In the remainder of this section, we will present statistics and detailed examples to highlight differences between our event annotations and the ACE event annotations.

The statistics we present are based on 62 documents from the ACE-2005 corpus that one of our annotators also annotated.⁴ Since our event types and subtypes are not directly comparable to the

⁴Due to space constraints we do not give statistics for the other annotator.

ACE event types, we say there is a “match” when both the following conditions hold:

1. The span of text that represents an event in the ACE event annotations overlap with ours.
2. The entities participating in the ACE event are same as the entities participating in our event.⁵

Our annotator recorded a total of 212 events in 62 documents. We found a total of 63 candidate ACE events that had at least two Person entities involved. Out of these 63 candidate events, 54 match both the aforementioned conditions and hence our annotations. A classification of all of the events (those found by our annotators and the ACE events involving at least two persons) into our social event categories and into the ACE categories is given in Figure 5. The figure shows that the majority of social events that match the ACE events are of type INR.VERBAL-NEAR. On analysis, we found that most of these correspond to the ACE type/subtype CONTACT.MEET. It should be noted, however, our type/subtype INR.VERBAL-NEAR has a broader definition than ACE type/subtype CONTACT.MEET, as will become apparent later in this section. In the following, we discuss the 9 ACE events that are not social events, and then we discuss the 158 social events that are not ACE events.

Out of the nine candidate ACE events which did not match our social event annotation, we found five are our annotation errors, i.e. when we analyzed manually and looked for ACE events that did not correspond to our annotations, we found that our annotator missed these events. The remaining four, in contrast, are useful for our discussion because they highlight the differences in ACE and our annotation perspectives. This will become clearer with the following example:

- (6) In central Baghdad, [a Reuters cameraman] and [a cameraman for Spain’s Telecinco] died when an American tank fired on the Palestine Hotel

ACE has annotated the above example as an event of type CONFLICT-ATTACK in which there are two entities that are of type person: the **Reuters cameraman** and the **cameraman for**

⁵Recall that our event annotations are between exactly two entities of type PER.Individual or PER.Group.

Spain’s Telecinco, both of which are arguments of type “Victim”. Being an event that has two person entities involved makes the above sentence a valid candidate (or potential) ACE event that we match with our annotations. However, it fails to match our annotations, since we do not annotate an event in this sentence. The reason is that this example does not reveal the cognitive states of the two entities – we do not know whether one was aware of the other.

We now discuss social events that are not ACE events. From Figure 5 we see that most of the events that did not overlap with ACE event annotations were Cognition (COG) social events. In the following, our annotator records a COG relation between **Digvijay Singh** and **Abdul Kalam** (also **Atal Behari Vajpayee** and **Varuna**). The reason is that by virtue of talking about the two entities, **Digvijay Singh’s** cognitive state contains those entities. However, the sentence does not reveal the cognitive states of the other two entities and therefore it is not an INR event. In contrast, ACE does not have any event annotation for this sentence.

- (7) The Times of India newspaper quoted [Digvijay Singh] as {saying} that [Prime Minister Atal Behari Vajpayee] and [President Abdul Kalam] had offended [the Hindu rain God Varuna] by remaining bachelors. **COG**

It is easy to see why COG relations are not usually annotated as ACE events. But it is counter-intuitive for INR social events not to be annotated as ACE events. We explain this using Example (3) in Section 2. Our annotator recorded an INR relation between **Toujan Faisal** and the **committee** (event span: *informed*). ACE did not record any event between the two entities.⁶ This example highlights the difference between our definition of Interaction events and ACE’s definition of Contact events. For this reason, in Figure 5, 51 of our INR relations do not overlap with ACE event categories.

4 Annotation Procedure

We used Callisto (a configurable workbench) (Day et al., 2004) to annotate the ACE-2005 corpus for

⁶The ACE event annotated in the sentence is of type “Personell-Elect” (span *election*) which is not recorded as an event between two or more entities and is not relevant here.

62 Documents			Conflict (5)	Contact (32)		Justice-* (13)	Life (7)			Transaction (2)	Not Found
			Attack	Meet	Phone-Write		Die	Divorce	Injure	Transfer-Money	
INR	Verbal	Near (66)	0	26	0	9	0	0	0	0	31
		Far (17)	0	0	3	3	0	1	0	0	10
	NonVerbal	Near (14)	3	0	0	0	2	0	0	1	8
		Far (3)	0	0	0	0	0	0	0	1	2
COG (109)			2	0	0	0	1	0	0	0	106
PPR (2)			0	0	0	0	1	0	1	0	0
PCR (1)			0	0	0	0	0	0	0	0	1
Errors			0	3	0	1	1	0	0	0	

Figure 5: This table maps the type and subtype of ACE events to our types and subtypes of social events. The columns have ACE event types and sub-types. The rows represent our social event types and sub-types. The last column is the number of our events that are not annotated as ACE events. The last row has the number of social events that our annotator missed but are ACE events.

the social events we defined earlier. The ACE-2005 corpus has already been annotated for entities as part of the ACE effort. The entity annotation is therefore not part of this annotation effort. We hired two annotators. Annotators opened ACE-2005 files one by one in Callisto. They could see the whole document at one time (top screen of Figure 6) with entities highlighted in blue (bottom screen of Figure 6). These entities were only of type PER.Individual and PER.Group and belonged to class SPC. All other ACE entity annotations were removed. The annotators were required to read the whole document (not just the part that has entities) and record a social event span (highlighted in dark blue in Figure 6), social event type, subtype and the two participating entities in the event.

The span of a event mention is the *minimum* span of text that *best* represents the presence of the type of event being recorded. It can also be viewed as the span of text that evokes the type of event being recorded. The span may be a word, a phrase or the whole sentence. For example, the span in Example (4) in Section 2 includes *strapped to his body* because that confirms the physical proximity of the two entities. We have, however, not paid much attention to the annotation of the span, and will not report inter-annotator agreement on this part of the annotation. The reason for this is that we are interested in annotating the underlying semantics; we will use machine learning to find the linguistics clues to each type of social event, rather than relying on the annotators' ability to determine these. Also note that we did not give precise

instructions on which entity mentions to choose in case of multiple mentions of the same entity. Again, this is because we are interested in annotating the underlying semantics, and we will rely on later analysis to determine which mentions participate in signaling the annotated social events.

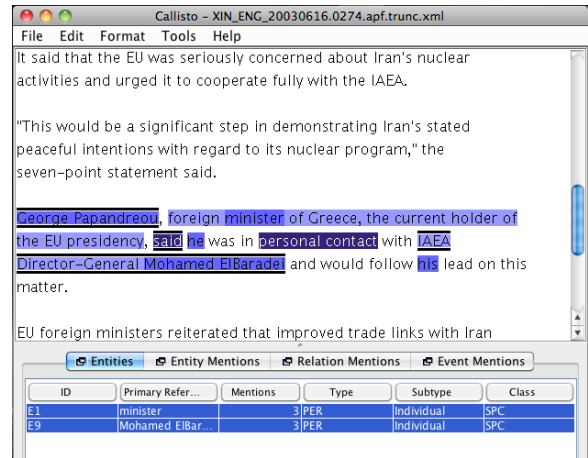


Figure 6: Snapshot of Callisto. Top screen has the text from a document. Bottom screen has tabs for Entities, Entity Mentions etc. An annotator selected text *said*, highlighted in dark blue, as an event of type COG between Entities with entity ID E1 and E9.

Both our annotators annotated 46 common documents. Out these, there was one document that had no entity annotations, implying no social event annotation. The average number of entities in the remaining 45 documents was 6.82 per document, and the average number of entity mentions per document was 23.78. The average number of social events annotated per document by one anno-

tator was 3.43, whereas for the other annotator it was 3.69. In the next section we present our inter-annotator agreement calculations for these 45 documents.

5 Inter-annotator Agreement

Annotators consider all sentences that contain at least two person entities (individuals or group), but do not always consider all possible labels, or annotation values. As represented in the decision tree in Figure 5, many of the labels are conditional. At each next depth of the tree, the number of instances can become considerably pruned. Due to the novelty of the annotation task, and the conditional nature of the labels, we want to assess the reliability of the annotation of each decision point. For this, we report Cohen’s Kappa (Cohen, 1960) for each independent decision. We use the standard formula for Cohen’s Kappa given by:

$$Kappa = \frac{P(a) - P(e)}{1 - P(e)}$$

where $P(a)$ is probability of agreement and $P(e)$ is probability of chance agreement. These probabilities can be calculated from the confusion matrix represented as follows:

	Yes _{Ann₂}	No _{Ann₂}
Yes _{Ann₁}	A	B
No _{Ann₁}	C	D

In addition, we present the confusion matrix for each decision point to show the absolute number of cases considered, and F-measure to show the proportion of cases agreed upon. For most decision points, the Kappa scores are at or above the 0.67 threshold recommended by Krippendorff (1980) with F-measures above 0.90. Where Kappa is low, F-measure remains high. As discussed below, we conclude that the annotation schema is reliable.

We note that in the ACE annotation effort, inter-annotator agreement (IAA) was measured by a single number, but this number did not take chance agreement into account: it simply used the evaluation metric to compare systems against a gold standard. Furthermore, this metric is composed of distinct parts which were weighted in accordance with research goals from year to year, meaning that the results of applying the metric changed from year to year. We have also performed an

ACE-style IAA evaluation, which we report at the end of this section.

Figure 7 shows the results for the seven binary decision points, considered separately. The number of the decision point in the table corresponds to the decision points in Figure 4. The (flattened) confusion matrices in column two present annotator two’s choices by annotator one’s, with positive agreement in the upper left (cell A) and negative agreement in the lower right (cell D). In all cases the cell values on the agreement diagonal (A, D) are much higher than the cells for disagreement (B, C). The upper left cell (A) of the matrix for decision 1 represents the positive agreements on the presence of a social event (N=133), and these are the cases considered for decision 2. For the remaining decisions, agreement is always unbalanced towards agreement on the positive cases, with few negative cases. In the case of decision 4, for example, this reflects the inherent unlikelihood of the NONVERBAL-FAR event. In other cases, it reflects a property of the genre. For example, when we apply this annotation schema to fiction, we find a much higher frequency of physically proximate events (PPR), corresponding to the lower left cell (D) of the confusion matrix for decision 6.

For decision 4 (NONVERBAL-NEAR) and 7 (PCR/COG), kappa scores are low but the confusion matrices and high F-measures demonstrate that the absolute agreement is very high. Kappa measures the amount of agreement that would not have occurred by chance, with values in [-1,1]. For binary data and two annotators, values of -1 can occur, indicating that the annotators have perfectly non-random disagreements. The probability of an annotation value is estimated by its frequency in the data (the marginals of the confusion matrix). It does not measure the actual amount of agreement among annotators, as illustrated by the rows for decisions 4 and 7. Because NONVERBAL-FAR is chosen so rarely by either annotator (never by annotator 2), the likelihood that both annotators will agree on NONVERBAL-NEAR is close to one. In this case, there is little room for agreement above chance, hence the Kappa score of zero. We should point out, however, that this skewness was revealed from the annotated corpus. We did not bias our annotators to look for a particular type of relation.

The five cases of high Kappa and high F-

measure indicate aspects of the annotation where annotators generally agree, and where the agreement is unlikely to be accidental. We conclude that these aspects of the annotation can be carried out reliably as independent decisions. The two cases of low Kappa and high F-measure indicate aspects of the annotation where, for this data, there is relatively little opportunity for disagreement.

Decision Point	Confusion Matrix				Kappa	F1
	A	B	C	D		
1 (+/- Relation)	133	31	34	245	0.68	0.80
2 (1 or 2 way)	51	8	1	73	0.86	0.91
3 (Verbal/NonV)	40	4	0	7	0.73	0.95
4 (NonV-Near/Far)	6	0	1	0	0.00	0.92
5 (Verbal-Near/Far)	30	1	2	7	0.77	0.95
6 (+/- PPR)	71	0	1	1	0.66	0.99
7 (PCR/COG)	69	1	1	0	-0.01	0.98

Figure 7: This table presents the Inter-annotator agreement measures. Column 1 is the decision point corresponding to the decision tree. Column 2 represents a flattened confusion matrix where A corresponds to top left corner, D corresponds to the bottom right corner, B corresponds to top right corner and C corresponds to the bottom left corner of the confusion matrix. We present values for Cohen’s Kappa in column 3 and F-measure in the last column.

Now, we present a measure of % agreement for our annotators by using the ACE evaluation scheme.⁷ We considered one annotator to be the gold standard and the other to be a system being evaluated against the gold standard. For the calculation of this measure we first take the union of all event spans. As in the ACE evaluation scheme, we associate penalties with each wrong decision annotators take about the entities participating in an event, type and sub-type of an event. Since these penalties are not public, we assign our own penalties. We choose penalties that are not biased towards any particular event type or subtype. We decide the penalty based on the number of options an annotator has to consider before taking a certain decision. For example, we assign a penalty of 0.5 if one annotator records an event which the other annotator does not. If annotators disagree on the relation type, the penalty is 0.25 because there are four options to select from (INR, COG, PPR, PCR). Similarly, we assign a penalty of 0.2

⁷<http://www.itl.nist.gov/iad/mig/tests/ace/2007/doc/ace07-evalplan.v1.3a.pdf>

if the annotators disagree on the relation sub-types (VERBAL-NEAR, VERBAL-FAR, NONVERBAL-NEAR, NONVERBAL-FAR, No sub-type). We assign a penalty of 0.5 if the annotators disagree on the participating entities (incorporating the directionality in directed relations). Using these penalties, we get % agreement of 69.74%. This is a high agreement rate as compared to that of ACE’s event annotation, which was reported to be 31.5% at the ACE 2005 meeting.

6 Conclusion and Future Work

We have presented a new annotation scheme for extracting social networks from text. We have argued, social network created by the sender - receiver links in Enron Email corpus can benefit from social event links extracted from the content of emails where people talk about their “implicit” social relations. Our annotation task is novel in that we are interested in the cognitive states of people: who is aware of interacting with whom, and who is aware of whom without interacting. Though the task requires detection of events followed by conditional classification of events into four types and subtypes, we achieve high Kappa (0.66-0.86) and F-measure (0.8-0.9). We also achieve a high global agreement of 69.74% which is inspired by Automated Content Extraction (ACE) inter-annotator agreement measure. These measures indicate that our annotations are reliable.

In future work, we will apply our annotation effort to other genres, including fiction, and to text from which larger social networks can be extracted, such as extended journalistic reporting about a group of people.

Please contact the second author of the paper about the availability of the corpus.

Acknowledgments

This work was funded by NSF grant IIS-0713548. We thank Dr. David Day for help with adapting the annotation interface (Callisto) to our requirements. We would like to thank David Elson for extremely useful discussions and feedback on the annotation manual and the inter-annotator calculation scheme. We would also like to thank the Natural Language Processing group at Columbia University for their feedback on our classification of social events.

References

- Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.
- David Day, Chad McHenry, Robyn Kozierok, and Laurel Riek. 2004. Callisto: A configurable annotation workbench. *International Conference on Language Resources and Evaluation*.
- Jana Diesner, Terrill L Frantz, and Kathleen M Carley. 2005. Communication networks from the enron email corpus it’s always about the people. enron is no different. *Computational & Mathematical Organization Theory*, 11(3):201–228.
- G Doddington, A Mitchell, M Przybocki, L Ramshaw, S Strassel, and R Weischedel. 2004. The automatic content extraction (ace) program—tasks, data, and evaluation. *LREC*, pages 837–840.
- P S Keila and D B Skillicorn. 2005. Structure in the enron email dataset. *Computational & Mathematical Organization Theory*, 11 (3):183–199.
- Bryan Klimt and Yiming Yang. 2004. Introducing the enron corpus. *In First Conference on Email and Anti-Spam (CEAS)*.
- Klaus Krippendorff. 1980. *Content Analysis: An Introduction to Its Methodology*. Sage Publications.
- Andrew McCallum, Xuerui Wang, and Andres Corrada-Emmanuel. 2007. Topic and role discovery in social networks with experiments on enron and academic email. *Journal of Artificial Intelligence Research*, 30 (1):249–272.
- Nishith Pathak, Colin DeLong, Arindam Banerjee, and Kendric Erickson. 2008. Social topic models for community extraction. *Proceedings of SNA-KDD*.
- Ryan Rowe, German Creamer, Shlomo Hershkop, and Salvatore J Stolfo. 2007. Automated social hierarchy detection through email network analysis. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 109–117.
- Stanley Wasserman and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.