Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media

Toward Social Causality: An Analysis of Interpersonal Relationships in Online Blogs and Forums

Roxana Girju
University of Illinois at Urbana-Champaign
Urbana, IL 61801
email: girju@illinois.edu

Abstract

In this paper we present encouraging preliminary results into the problem of `social causality` (causal reasoning used by intelligent agents in a social environment) in online social interactions based on a model of reciprocity. At every level, social relationships are guided by the shared understanding that most actions call for appropriate reactions, and that inappropriate reactions require management. Thus, we present an analysis of interpersonal relationships in English reciprocal contexts. Specifically, we rely here on a large and recently built database of 10,882 reciprocal relation instances in online media. The resource is analyzed along a set of novel and important dimensions: *symmetry, affective value, gender*, and *intentionality of action* which are highly interconnected. At a larger level, we automatically generate *chains of causal relations* between verbs indicating interpersonal relationships. Statistics along these dimensions give insights into people's behavior, judgments, and thus their social interactions.

Causal knowledge and reasoning are central components of any inference system. Identifying the actual causes of events will help us gain a better understanding of the human language and of the world. Unfortunately, physical causes and effects (about processes and natural phenomena) (Pearl 2000; Hobbs 2001; Girju 2003) are simply inadequate for exploiting and explaining social phenomena. In contrast, `social causality` (causal reasoning used by intelligent agents in a social environment) is qualitatively different from physical causal reasoning. For example, social causality emphasizes multiple causal dimensions such as judgments of intention, responsibility, credit or blame, and other factors people use in reasoning about social events and other people's actions (Malle and Bennett 2002). According to cognitive science theories, such judgments are fundamental social explanations (Shaver 1985).

Modeling social causality and inference can guide conversation strategies, facilitate the modeling and understanding of social emotions, and bring new insights into issues such as explanations, intentions, and perception. Moreover, this modeling is very important for the creation of user-aware adaptive interfaces and systems that socially interact with and for people.

We are particularly interested in social causality as captured through an analysis of interpersonal relations in online social media (e.g., blogs and forums). Specifically, we would like to determine how people treat or reciprocate each other. At every level, social relationships are guided by the shared understanding that most actions call for appropriate reactions, and that inappropriate reactions require management. According to sociologists and philosophers, the concept of reciprocity lies at the foundation of social organization as it strengthens and maintains social relations among people. The way people perceive other people's actions and behavior and the way these are expressed in language play an important role in governing people's behavior, judgments, and thus their social interactions.

The model can also be successfully applied to specific datasets – i.e., specific online groups or communities, where it can identify, monitor across time, analyze, and even predict the behavior of specific people or organizations (e.g., to monitor cooperation, trustworthiness and personality). Moreover, the model can also facilitate the comparison of interpersonal relationships across cultures (e.g., ethnicity in social networks).

In this paper we present an analysis of interpersonal relationships encoded by a very pervasive set of English reciprocal contexts and present a way to model them, along with insightful observations. Specifically, we rely here on a large and recently built database of 10,882 reciprocal relation instances in online media (Paul, Girju, and Li 2009). The resource is analyzed along a set of novel and important dimensions: *symmetry, affective value, gender*, and *intentionality* which are highly interconnected. At a larger level, we automatically generate *reciprocal (causal) chains* between verbs indicating interpersonal relationships. Statistics along these dimensions give insights into people's behavior, judgments, and thus their social interactions.

The paper is organized as follows. In the first four sections we give a summary of the reciprocity model and the generated database (described in detail in (Paul, Girju, and Li 2009)). Specifically, in the first section we present the concept of reciprocity followed by relevant previous work. In Sections 3 and 4 we briefly present the model and the reciprocity database. Our contribution to this paper is presented in the remaining sections. In particular, in Section 4 we provide a detailed analysis of the resource along the set of newly identified dimensions. In Section 5 we present the experimental data and results. Discussion and conclusions are presented in Section 6.

Reciprocity in language

The Oxford English Dictionary Online[1] defines reciprocity as "*a state or relationship in which there is mutual action, influence, giving and taking, correspondence, etc., between two parties*", while in WordNet (Fellbaum 1998) the verb *to reciprocate* means "*to act, feel, or give mutually or in return*". Mathematically, reciprocity can be defined as a relation between two eventualities e_1 and e_2:

$$\Re(e_1(\text{X, Y}), e_2(\text{Z, W})),$$

where each eventuality is an event, action, or state between two entities (animate or inanimate) which are the arguments representing the subject and the object. Modeling reciprocity offers the possibility of capturing not only the two reciprocal eventualities e_1, e_2, but chains of such eventuality pairs, thus, allowing for an in-depth analysis of the social behavior of two or more people, and even groups and communities.

The eventuality arguments captured in the reciprocity relation $\Re$ may or may not be explicitly stated in the sentence, but can be inferred. Two types of reciprocal relations can be identified: a) direct reciprocity (the relation is such that $X = W$ and $Y = Z$), and indirect reciprocity (the relation is such that $X \neq W$ and $Y = Z$). In this research we will focus only on direct reciprocal relations of eventualities, where the arguments are people, groups or organizations.

From the timing point of view there are two distinct possibilities: mutually performed eventualities that occur concurrently[2], written as $e_1(X, Y)$ & $e_2(Z, W)$, and "in return" reciprocity when one eventuality causes (reciprocates) the other, written as $e_1(X, Y) \prec_c e_2(Z, W)$.

A few such examples are presented below with the corresponding reciprocity relationships:

(1) Mary **argued with** Paul at the station.
argue_with(Mary, Paul) & argue_with(Paul Mary)

(2) Paul and Mary hate **each other**.
hate(Paul, Mary) & hate(Mary, Paul)

(3) Mary likes Paul **and** he likes her, **too**.
like(Mary, Paul) & like(Paul, Mary)

(4) Mary likes Paul **for** helping her sister.
help(Paul, Mary's sister) $\prec_c$ like(Mary,Paul)[3]

As shown in the examples above, in English there are two basic types of reciprocal constructions: mono-clausal reciprocals (involving words such as *(to) hug, to agree/argue with, partner of, mutual(ly), together, each other* – examples (1) and (2)) or sentence-level reciprocals (involving two consecutive clauses – examples (3) and (4)). Most of the sentence-level reciprocals are paraphrased by coordinations or subordinations of two clauses with the same or different predicate and most of the time inverted arguments. They might also manifest various markers (as shown in bold).

So far we have focused only on sentence-level contexts when the eventualities occur in different consecutive clauses, and when the subject – object arguments of each eventuality are personal pronoun pairs which occur in reverse order in each eventuality. One such example is "**She** *likes* **him** *for helping* **her**". Here the two eventualities are *like(she, he)* and *help(he, she)*. In this example, although the subject of the second verb is not explicitely stated, it is easily inferred. These simplifying assumptions proved very useful in reciprocity discovery procedure to ensure the accuracy of the discovered patterns and their matched instances (Paul, Girju, and Li 2009) as shown briefly in Section 3.

Previous work

Social causality has been studied by social scientists to better understand human social interactions and to develop more effective social intervention strategies. The concept of reciprocity has also been studied a lot in different disciplines such as social sciences (Gergen, Greenberg, and Willis 1980), anthropology (Sahlins 1972), and philosophy (Becker 1990) – but mainly from the perspective of ethics.

In linguistics, this problem has been tackled only more recently. Most of the work on reciprocity focuses on quantifiers like *each other* and *one another* (Dalrymple et al. 1998; Heim 1991; König 2005) and has been done by language typologists (Maslova and Nedjalkov 2005; Haspelmath 2007) who are interested in how reciprocal constructions of this type vary across large sets of world's languages.

Moreover, social dynamics has been studied a lot in social networks which combine network topology with computational models applied mostly on data from on-line networks (i.e., who talks to whom, the time and frequency of interaction – but not based on *what* is said and meant). Although this approach may describe some typical patterns, it provides limited insight into human social interactions which are captured most of the time through language.

To the best of our knowledge, in computational linguistics and information systems (i.e., social and information networks) the problems of modeling and analyzing social causality and reciprocity in language are novel. Specifically, our approach provides a social interaction framework that explains not only *who* and *how* the identified people interact with each other, but also helps one understand *why* an event was performed in response to another.

In the next sections we give a brief description of the reciprocity model and the generated database.

Reciprocity model

Our algorithm first discovers clusters of patterns indicating reciprocity in English, and then merges the resulting clusters to identify the final set of reciprocal constructions.

So far we have focused on reciprocal eventualities which occur in two consecutive clauses and have two arguments:

[1] http://www.oed.com/

[2] The word "concurrently" also refers to cases like "*John and Mary chase each other*" – a mutual iterative process.

[3] We assume here that the subject of the verb *help* has been recovered and the coreference solved.

a subject and an object. One way to identify this information is to fully parse each sentence of a corpus and identify coordinations or subordinations of two clauses. Then locate the subject and object arguments of each verb in each clause with the help of a PropBank-style grammatical or semantic role labeler (Kingsbury, Palmer, and Marcus 2002) making sure they represent people named entities (as indicated by proper names, personal pronouns, etc.). Since our focus is on reciprocal constructions, the verbs have to have the same set of arguments (subject-object) in reverse order. Thus, noun and pronoun coreference should also be resolved at this point.

Since this procedure is rather complex and error-prone, the current version of our algorithm relies on a set of pronoun templates, where personal pronouns are anchor words (they have to be matched as such). Each template consists of four personal pronouns corresponding to a subject - object pair in one clause, and a subject - object pair in the other clause. Two such examples are

"[Part1] **I** [Part2] **him** [Part3] **he** [Part4] **me** [Part5]" and

"[Part1] **they** [Part2] **us** [Part3] **we** [Part4] **them** [Part5]",

where [Part1] - [Part5] are partitions identifying any sequence of words. This is an elegant procedure since in English, pronouns have different cases such as nominative and accusative[4] which identify the subject, and respectively the object of an event. In English, there are 30 possible arrangements of nominative - accusative case personal pronoun pairs. Thus we built 30 pronoun templates.

Since the first two pronouns in each pronoun template belong to the first clause (C1), and the last two to the second clause (C2), the templates can be restated as

[Part1] C1 [Part3] C2 [Part5],

with the restriction that partition 3 should not contain any of the four pronouns in the template. C1 denotes "Pronoun1 [Part2] Pronoun2" and C2 denotes "Pronoun3 [Part4] Pronoun4". Partitions 2 and 4 contain the verb phrases (and thus the eventualities) we would like to extract. For speed and memory reasons, we limited their size to no more than 5 words.

Moreover, since the two clauses are consecutive, we hypothesize that they should be very close to each other. Thus, we restrict the size of each partition 1, 3, and 5 to no more than 5 words. We then considered all possible variations of the pattern where the size of each partition varies from 0 to 5 (216 possible combinations). Moreover, to ensure the accuracy of the procedure, partitions 1 and 5 are bounded to the left and respectively to the right by punctuation marks, parentheses, or paragraph boundaries. An example of an instance matched by one such pattern is "*,* **I** *cooked dinner for* **her** *and* **she** *loves* **me** *for that* ."

We searched the 30 pronoun templates with various partition sizes on a 20 million word English corpus obtained from Project Gutenberg, the largest single collection of free electronic books (over 27,000) (http://www.gutenberg.org) and British National Corpus (BNC), an 100 million word collection of English from spoken and written sources. We identified 1,613 distinct types of patterns which generated 1,866 distinct pattern instances. Thus, we selected the top 15 patterns (Table 1, column 1), after manual validation. These patterns represented 56% of the data. All the other patterns were discarded as having very low frequencies and being very specific.

The pattern discovery procedure yields an accuracy of 97% which shows that these patterns are highly accurate indicators of reciprocity in English. A detailed description and evaluation of this procedure is presented in (Paul, Girju, and Li 2009).

Reciprocity database

Representing the data. Using the 15 identified patterns we extracted pairs of eventualities of the form (e_1, e_2). This involves both reducing the clauses into a form that is semantically representative of some eventuality, as well as determining the order of the two eventualities.

As shown in the previous sections, each pattern contains two clauses of the form "$Pronoun_i$ [Part2/4] $Pronoun_j$", where the first pronoun is the subject and the second is the object. From each clause we extract the verb which was first stemmed and then negated if it was preceded by *not* or *n't*. For example, "*They do not like him because he snubbed them*" is represented as the eventualities $(e_1, e_2) = (snub, \neg like)$.

For each verb we also capture additional information such as tense and modality.

Data collection. While the Gutenberg and BNC collections are useful in obtaining the frequent patterns, they do not contain a very large number of eventuality pairs to do meaningful clustering. We thus queried the web through Google to easily obtain thousands of examples. For this we relied on the top 15 patterns and all pronoun combinations thereof (e.g. "*they * us because we * them*") and took the top 500 results for each pattern/pronoun combination (15*30*500)[5]. We then extracted the clauses from the result snippets using the procedure outlined in the previous section and obtained 10,882 (4,403 unique) reciprocity pairs since some of the queries had less than 500 matched instances[6].

A closer analysis of the generated database showed that the reciprocity instances were collected mostly from online blogs, forums, and message boards (96% of a set of 50 randomly selected instances were extracted from such online social media). For example, the instance sentence "*I only married you because you forced me.*" was extracted from Purse Forum at $http://forum.purseblog.com/relationships-and-family/im-so-torn-please-need-your-advice-381681.html$.

In the next sections we introduce a set of novel and important social dimensions along which this dataset can be ana-

[4]In English, the pronouns *you* has the same form in nominative and accusative.

[5]This is because Google limits traffic. However, in the future we can acquire more instances.

[6]The reciprocity dataset is available for download at http://apfel.ai.uiuc.edu/resources.html.

lyzed and present various experiments whose results bring new insights into social reciprocity and interpersonal relationships in online reciprocal contexts.

Properties of interpersonal verbs and reciprocal instances

The reciprocity model has identified verbs and verb pairs as important source of information for detecting reciprocity. In this section we introduce a set of attributes of verbs and reciprocal instances which are important in the analysis of social reciprocal interactions: *symmetry, affective value, gender*, and *intentionality of actions*. These are extended with grammatical features like tense, aspect, modality, voice, transitivity and negation. We analyze the reciprocity dataset along these dimensions which are highly interconnected. They are presented next.

A. Symmetry is a useful property for identifying the order of the eventualities in reciprocal relationships and can be very useful in building reciprocal chains. The reciprocal relation between two eventualities can be symmetric or asymmetric. The relation is symmetric when e_1 & e_2 is true (e.g., "*He loves her* **and** *she loves him*"), and asymmetric when one eventuality is performed "in return" for the other (e.g., "*He ignores her* **when** *she scolds him*"). There are, however ambiguous situations as well – i.e., those contexts that can encode both symmetric and asymmetric instances (e.g., *He loves her* **and** *she loves him*" and "*He cheated on her* **and** *she still loves him!*").

Based on our quick check of the patterns discovered and based on the previous linguistic literature on clause-level reciprocal constructions (König 2005), we have identified a list of 6 linguistic features which are good indicators of symmetric reciprocal relationships in sentence-level reciprocal instances. Table 1 provides an overview of the symmetry of reciprocity for each of the 15 patterns considered along with examples. Columns 1, 3, 4, and 5 show the features which help determine if the reciprocity relationship encoded by a pattern instance is symmetric or not. These features are presented next. We use the notations e_1 and e_2 to denote the two eventualities, where the index represents the order in which they are mentioned in context.

`F1. Reciprocal context` (col. 1). This feature indicates one of the 15 patterns identified. Some patterns indicate sequential eventualities, and thus impose an asymmetric reading. For example, "*She married him because he made her laugh*" shows an asymmetric reciprocity, while the reciprocity in "*I love him as much as he loves me*" is symmetric.

`F2, F3. Type of eventuality` indicates if eventualities e_1 (col. 3) and e_2 (col. 4) are states or events. For example, verbs describing states refer to the way things "are" – their appearance, state of being, smell, etc. (e.g., *need, hate, love*), while *hit and chase* are action verbs. The values of these features are automatically determined based on an in-house list of 300 stative verbs identified from WordNet. The identification procedure captured the most important difference between stative and action verbs in that action verbs can be used in continuous tenses while stative verbs can not.

This feature was borrowed from the linguistic literature on clause-level constructions such as "each other" (König 2005). König, for example suggests that with predicates denoting states the relevant sentences express fully symmetric situations (e.g., "*These two hate each other.*"), whereas event denoting predicates are more compatible in their interpretation with a delay between the two relevant events (e.g., "*They chase each other.*").

We hypothesize here that these observations can be extended to sentence-level reciprocal constructions between distinct verbs as well. Moreover, we show that in our dataset this delay between the two events in asymmetric instances corresponds to "in return" reciprocity (social causality).

`F4 and F5. Verb modality` represents the modality of each verb (if any). Possible values are: may, would, can, shall, might, will, could, should, must.

`F6. Relative temporal order of the two eventualities` (col. 5). This feature indicates if (and if yes, which) one eventuality happens before, after, or in the same time with the other eventuality. The order is simply calculated based on the tense information provided by each verb in context. For example, past simple happens before present or future tense. This feature is used to further "disambiguate" those instances for which the symmetry property can not be determined solely based on the pattern information (feature F1).

For example, for the pattern "C1 and C2", if the eventualities e_1 and e_2 are states (or events), then the encoded reciprocity relation is symmetric (or asymmetric respectively). When the pattern is "C1 as much as C2", if the eventualities are events and the eventuality e_2 happens before e_1, then the encoded reciprocity relation is asymmetric.

In order to implement the features we first chunk parsed (Li and Roth 2001) each pattern instance and automatically identified the verbs and their tense information. In Table 1 we provide a summary of the feature values identified. For eventuality type we use the term "mixed" to refer to a either a state or an event. The symmetry column was filled out manually by looking at the identified feature values and considering some examples.

Table 1 indicates that the patterns "C1 and C2 back", "C1 when C2", "C1 whenever C2", "C1 because C2", "C1 for C2 (vb-ing)", "C1 for what C2", "C1 and thus C2", "when C1, C2", and "C1 as long as C2" are asymmetric irrespective of the type of the two eventualities. The analysis indicates that all the other patterns can be either symmetric, or asymmetric if their eventualities are either states or events, respectively. For these last patterns, when the eventuality type is mixed, the relative temporal order of the verbs identifies the order of the eventualities.

The reciprocity dataset contains 5,880 ambiguous examples – that is, examples from one of the four predetermined ambiguous patterns: *C1 and/but/[.;,]/as much as C2*. From these examples, 300 random examples were set aside for testing and the rest were used to train the classifier. 24 of these instances did not encode reciprocity and were thus eliminated. In the remaining set the order of the eventualities was identified by two judges who agreed on all the

Patterns	Freq.	Event. type		Rel. temporal order of event.	Symmetry	Examples
		e_1	e_2			
C1 [, |; |.] C2	10.45%	state	state	$e_1 =_t e_2$	symmetric	*He loves her; she loves him.*
		event	event	$e_1 =_t e_2$	asymmetric $e_2 \rightarrow e_1$	*He helped me, I helped him.*
C1 and C2	27.03%	state	state	$e_1 =_t e_2$	symmetric	*They respect him and he respects them.*
		event	event	$e_1 =_t e_2$	asymmetric $e_2 \rightarrow e_1$	*He hugs her and she elbows him.*
C1 and C2 back	0.73%	state	state	$e_1 =_t e_2$	symmetric	*She does love him and he loves her back.*
		event	event	$e_1 =_t e_2$	asymmetric $e_2 \rightarrow e_1$	*She kissed him and he kissed her back.*
C1 and C2 for that	0.055%	mixed	mixed	$e_1 \leq_t e_2$	asymmetric $e_2 \rightarrow e_1$	*He destroyed her life and she hates his for that.*
C1 and C2, too	0.21%	state	state	$e_1 =_t e_2$	symmetric	*He loves her and she loves him.*
		event	event	$e_1 =_t e_2$	asymmetric $(e_2 \rightarrow e_1)$	*I chase him and he chases me, too.*
C1 when C2	8.82%	mixed	mixed	$e_1 \geq_t e_2$	asymmetric $(e_1 \rightarrow e_2)$	*He ignores her when she scolds him.*
C1 whenever C2	1.71%	mixed	mixed	$e_1 \geq_t e_2$	asymmetric $(e_1 \rightarrow e_2)$	*He was there for her when she needed him.*
C1 because C2	11.45%	mixed	mixed	$e_1 \geq_t e_2$	asymmetric $(e_1 \rightarrow e_2)$	*She married him because he was good to her.*
C1 as much as C2	16.55%	state	state	$e_1 =_t e_2$	symmetric	*She enjoyed him as much as he enjoyed her.*
		event	event	$e_1 \geq_t e_2$	asymmetric $(e_1 \rightarrow e_2)$	*They hit him as much as he hit them.*
C1 for C2 (vb-ing)	7.77%	mixed	mixed	$e_1 >_t e_2$	asymmetric $(e_1 \rightarrow e_2)$	*They thanked him for helping them.*
C1 but C2	0.10%	state	state	$e_1 =_t e_2$	symmetric	*I love her but she hates me.*
		mixed	mixed	$e_1 \geq_t e_2$	asymmetric $(e_2 \rightarrow e_1)$	*He tried to talk to her but she ignores him.*
C1 for what C2	2.48%	mixed	event	$e_1 \geq_t e_2$	asymmetric $(e_1 \rightarrow e_2)$	*They will punish him* **for** *what he did to them.*
C1 and thus C2	0.046%	mixed	event	$e_1 \leq_t e_2$	asymmetric $(e_2 \rightarrow e_1)$	*She rejected him and thus he killed her.*
when C1, C2	10.91%	mixed	mixed	$e_2 \geq_t e_1$	asymmetric $(e_2 \rightarrow e_1)$	*When he started to hit them, they arrested him.*
C1 as long as C2	1.69%	mixed	mixed	$e_2 \geq_t e_1$	asymmetric $(e_2 \rightarrow e_1)$	*She is staying with him as long as he is kind to her.*
Total	10,822					

Table 1: The set of 4 (out of 6) features indicative of symmetric reciprocity shown here along with examples. The "|"s in the pattern C1 [, |; |.] C2 mean "or". constructions.

instances.

We compared the performance of three learning algorithms: Perceptron, Winnow, and Naïve Bayes trained on the dataset using the identified feature set. The algorithm which yielded highest accuracy was the perceptron algorithm with learning rate 0.1 and threshold 4. It achieved an accuracy of 84.33% on the test set. 78.33% of the examples are false, which means our classifier reduces 37.1% of the error made by a guess-all-false baseline.

The results indicate that the four identified features prove to be very good indicators of the symmetry and eventuality order in reciprocal contexts.

B. Affective value. Since we are interested in analyzing how people interact, we also identify the affective value (polarity) associated with each eventuality in reciprocal contexts: Good, Bad, Neutral. So far (Paul, Girju, and Li 2009), we successfully used the Subjectivity Clues lexicon (Wilson, Wiebe, and Hoffmann 2005) which provides over 8,000 word entries. From 10,882 eventuality pairs in our database, 40.1% of the total number of verbs and 36.9% of the verb pairs were found in the subjectivity lexicon. The experiments showed that each polarity class is most likely to be reciprocated by itself: Good for Good (altruism) and Bad for Bad (retaliation), and that it is more likely that people respond with Good actions to Bad behavior ("turn the other cheek") than with Bad for Good.

Table 2 shows all possible combinations of pairs of affective values and their associated probabilities in the corpus. These values are computed for those pairs where both words have known polarity.

	Good	**Bad**	**Neutral**	**Total**
Good	0.90	0.18	0.29	0.63
Bad	0.09	0.82	0.08	0.29
Neutral	0.01	0.002	0.63	0.09

Table 2: All possible combinations of pairs of affective values and their associated probabilities as found in the corpus. The numbers show conditional probabilities P($row_i|col_j$). The Total column is the probability of each affective class (P(row_i)).

In this paper we extend our previous results with a 4-node Hidden Markov Model (HMM) which can model sequential data such that each piece of data is generated by some state, and the state of the next piece of the sequence depends on the current state. This is a natural approach for our data, since the reciprocated eventuality "follows" the original one (denoted here as e_r and e_o, respectively)[7]. (In the case of a symmetric relationship, we generated two directed links, one for each direction.). Thus, the nodes t_0 and t_3 belong to a designated "start/end class" (which we will call z_0), and t_1 and t_2 belong to some classes (≥ 1) from which e_o and e_r are respectively drawn.

In the HMM-based method we associate each class z with both a distribution of verbs and a distribution of affective values. Thus, in the generative process, after choosing a class z we independently sample a verb e from $P(e|z)$ and an affective value f from $P(f|z)$, which has a Dirichlet(δ) prior. We also introduce a parameter λ that determines the probability that f is actually drawn from $P(f|z)$, while $(1-\lambda)$ would be the probability that f is chosen at random. This helps account for noise in our subjectivity lexicon. For this procedure we used the Subjectivity Clues.

Thus, the probability of generating an affective value and an eventuality (f, e) is:

$$P(f,e) = \sum_{k}^{|Z|} P(z=k|z_{prev})(\lambda P(f|z=k) + (1-\lambda)\frac{1}{F})P(e|z=k) \quad (1)$$

where F is the number of possible affective values (in our case, 3).

The Gibbs sampling update equation is:

$$P(z_{io}=j, z_{ir}=k|e_{o,i}=a, e_{r,i}=b, f_{o,i}=c, f_{r,i}=d, \mathbf{z_{-i}}, \alpha, \beta) \propto$$
$$(n^{z_o=j}+\gamma) \times \frac{n^{z=j}_{e_o=a}+\beta}{n^{z=j}_{e_o=*}+V_o\beta} \times \frac{n^{z_r=k}+\gamma}{n^{z_o=j}+C\gamma} \times \frac{n^{z=k}_{e_r=b}+\beta}{n^{z=k}_{e_r=*}+V_r\beta}$$
$$\times((1-\lambda)(\frac{1}{F})+\lambda\frac{n^{z=j}_{f=c}+\beta}{n^{z=j}_{f=*}+F\delta}) \times ((1-\lambda)(\frac{1}{F})+\lambda\frac{n^{z=k}_{f=d}+\beta}{n^{z=k}_{f=*}+F\delta}) \quad (2)$$

In the case that a verb's affective value is not known, we simply ignore this component and compute $P(z|e_o, e_r)$ (i.e., $\lambda = 0$ in these cases).

Thus we can predict the affective value of words whose value is unknown. Once the model is learned, we can compute $P(f|word)$ by marginalizing across all classes. That is, $P(f|word) = \sum_z^C P(f|z)P(z|word)$, where $P(z|word) = \frac{P(word|z)P(z)}{P(word)}$ by Bayes theorem.

We learned the model with 20 classes and $\lambda = 0.8$, then computed $P(f|word)$ for words that were not in the subjectivity lexicon. For each affective class, we took the 30 words with the highest value of $P(f_i|word)$, although some affective classes had fewer than 30 words assigned to them. The remaining 75 words were presented (in a random order) to two judges to rate as *Good*, *Bad*, or *Neutral*. The judges agreed on 51 of the words. Our affective value clustering can predict an unknown verb's affective value with 51% accuracy, while it can discriminate between *Good* and *Bad* with 68% accuracy.

C. Intentionality of actions. Some interpersonal verbs exhibit what is called an inherent intentionality. For example, actions such as *forgive, revenge, slay* are intentional irrespective of the reciprocal context in which they are used. The intentionality of others such as *hit* and *touch* is highly context dependent, while others seem to be unintentional (e.g., *love*). This property is very important for a model of reciprocity and thus social interactions, since the intentionality and the affective values of interpersonal verbs are highly correlated with blame and responsibility (Malle and Bennett 2002). For example, "*Mary kicks Paul and he kicks her*" shows an asymmetric reciprocal context entailing that his kicking is intentional. Moreover, since the verb predicates are the same, are asymmetric, and have negative affective values, then her kick is perceived as intentional as

[7]This is possible after we determine the order of the eventualities.

well. This observation is also supported by the behavioral psychology hypothesis (Knobe 2003) which states that when people evaluate negative actions they take intentionality into account, but fail to do so when they evaluate a positive action[8]. Thus, people's judgment of the intentionality of one's behavior is claimed to be significantly influenced by moral considerations. Intentionality and moral considerations also have an impact on people's causal judgments (Knobe 2003).

We will test this hypothesis inductively through a rule which takes into account affective value, responsibility, and intentionality in reciprocal contexts: if a reciprocal instance is generated by a symmetrically ambiguous pattern, and the two eventualities are actions whose affective values are bad, than the instance is asymmetric and e_o is intentional.

Analyzing social interactions

Once the dataset was annotated with the identified properties, we can now compute various statistics which will shed light on how people reciprocate. For example, statistics show that 54.34% of the dataset was encoded by ambiguous symmetric patterns. Overall, the top frequent verb pairs (15% of the data) are *need – need* and *love – love* followed by similar pairs *hate – hate, miss – miss, like – like, asked – told, know – follow, want – want*.

Gender analysis. Since we are also interested in gender-based statistics, we considered instances where each participant is male and female (e.g., "*he* vb1 *her because she* vb2 *him*"). 22.55% of the eventualities in unambiguous patterns were initiated by males, while only 12.49% by females. Statistics also show that men retaliate more often than women.

In general, the data show that men are more violent and aggressive, whereas women are more forgiving. This depends on the reciprocal class, however. Consider the cluster whose e_o words include *punish*, *refuse*, *criticize*, and *reject*. The top e_r words for men are *accept*, *hug*, *tolerate*, and *owe*. On the other hand, top e_r words for women include *cheat*, *dump*, and *despise*. It seems that men are more forgiving in the face of criticism and rejection, while women are more forgiving in response to embarrassment. Furthermore, it seems that men and women are generally mutually respectful; it is only when that respect is broken that their responses may differ (i.e., women respond to *cheat* with *hate, despise* and *sue*, while men prefer to *dump* and *divorce*).

Some verbs are more strongly associated with men - e.g., *rape* (73%), *attack* (70%), and *blame* (77%). Some verbs that are more common to men such as *hire* and *arrest* are likely due to the prominence of men in authoritative positions. The verbs *emasculate* and *babysit* were specific to the female distributions, while verbs like *nag* were more frequently associated with women. Statistics also show that men are *loved, adored, needed, respected*, but also *hated* and *slapped*. And in such situations they respond back with *love, adore, need, respect*, but also with *hate* for *hate*.

An interesting evaluation can also be done per pattern. For example, *because*, and *for -ing* instances explain why people behave in a particular way: in general we *love* people who *love/understand/care/need/entertain/complete/protect* us. We *hate* people who *hate/persecute/hurt/rob/beat* us. The dataset also shows that we *help* people who *need* us more often than we *leave* people who *need* us.

Another interesting observation is that in "for -ing" instances, 42% of the e_r's are "*thank*": people thank for *help, join, rescue, invite, support*, and *remember*.

Intentionality of actions. The definition of reciprocity identifies the e_r eventualities as always intentional (i.e., a response is always performed "on purpose"). However, e_o eventualities may or may not be intentional - and this may be difficult to detect especially in symmetrically ambiguous contexts. Thus, we evaluated the blame and responsibility hypothesis for the set of ambiguous pattern instances for a randomly selected subset of 100 ambiguous Bad-Bad instances and 100 ambiguous Good-Good instances. The hypothesis identified the e_o as intentional for 92% of the Bad-Bad instances and only 8% of the Good-Good ones with an accuracy of 87% against the gold standard generated by two annotators.

The dataset shows that in ambiguous instances, the bad actions initiated by women are perceived as intentional more often than those initiated by men. The top bad intentional actions initiated by women were *reject, dump, grab, hit, shout*, while those of men were *sue, fine* and *fight with*.

Reciprocity chains provide better insights into how people reciprocate. There are two types of such chains: dyadic and generic. Dyadic chains are formed between two people and are of the form "A $verb_1$ B → B $verb_2$ A → A $verb_3$ B . . ."[9], and are very useful in micro-level social interaction analysis. They can provide information about the social interaction behavior of a particular person or can identify how two specific persons (or people of different gender) treat each other. Generic chains are similar to the dyadic ones, but here the pairs of reciprocal instances are glued at the verb level and not at the argument (participant) level: "$verb_1 \rightarrow verb_2 \rightarrow verb_3 \rightarrow \ldots$". Since the chains abstract over the participants to these eventualities, they are very useful in macro-level social interaction analyses showing the behavior of a community or group as a whole. In this research we could generate only generic reciprocity chains (a total of 114,789) using only the "in return" reciprocal instances for which we identified the following types:

`1) Retaliation with increased magnitude.` These chains can end either in an end state (e.g., death (for *kill*)), or an affect state (e.g., *hate, blame, resent*):
criticize → insult → slap → stab → kill
criticize → insult → beat → dump → resent

`2) 'Good for Good' reciprocity chains:`
ask → help → thank

`3) 'Good for Bad' chains` end with a 'turn the other cheek' relationship, when one of the participants wants to end the retaliation chain:

[8] For example, "*The cartel destroyed the forest to increase its profits*", but not "*The cartel helped the forest to increase its profits*".

[9] The index of the verb denotes the order in which A and B reciprocate each other in text.

ignore → cheat → dump → sue → not blame

Overall, the 'Good for Good' chains are short (size 3-4 at most) compared with retaliatory or 'Good for Bad' ones which have a size of 6 on average (maximum 9). Moreover, it is interesting to note that for women the chains of retaliation with increased magnitude end most of the time in an affect state (e.g., *hate, resent*), while men end the fight with an action (e.g., *beat, divorce*).

An analysis of reciprocity chains per pattern shows that the "*for -ing*" and "*for what*" chains are much less noisy than "*because*", but also shorter. While most of them are retaliatory, some end with 'turn the other cheek' (when one party gives up). This is due to the fact that these patterns are rich in presupposition and entailment information: *forgive, criticize, punish*, etc.

Moreover, it seems that the chains generated per pattern are less noisier than computed overall – we obtained an accuracy of 78% (for a set of 50 randomly selected chains from the overall chain set) and respectively of 92% on average (for a set of 50 randomly selected chains per pattern). A quick analysis also shows that although most of these chains are highly accurate, their formation is not an easy process. For example, figurative meaning of such eventualities can infiltrate the literal context (e.g., "*He thanked them for embarrassing him*") – but we leave this for future research.

Discussion and Conclusions

In this paper we presented encouraging preliminary results into the problem of social causality in online social interactions based on a computational model of social reciprocity.

The experimental results provided nice insights into the problem, but can be further improved. For example, our model relies on web queries that capture reciprocal relationships in a limited context, thus failing to identify the full spectrum of interpersonal relationships existent on the web. One possible solution is to extend the model such that: (1) it identifies other named entities than pronouns, and (2) it considers a larger context (i.e., the two reciprocal eventualities are separated by larger text spans).

We also showed that this approach is a fruitful avenue for social interaction analysis. Interpersonal relations are the basis for social groups; they reflect emotions like affections, business interactions or some other social contacts. People in a relationship tend to influence each other, share ideas and engage in similar activities. Thus, the model can also be successfully applied to specific datasets – i.e., specific online groups or communities, where it can identify, monitor across time, analyze, predict the behavior of specific people or institutions, such as the government[10] or the police[11].

Moreover, we believe that such a resource of social causality can be very useful in a number of applications, ranging from question answering and textual entailment due to its rich semantic and pragmatic content.

[10] E.g., Buying political support and favors.

[11] E.g., "*The police responded to the protesters with tear gas*."

References

Becker, L., ed. 1990. *Reciprocity*. Chicago: University of Chicago Press.

Dalrymple, M.; Kazanawa, M.; Kim, Y.; Mchombo, S.; and Peters, S. 1998. Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy* 21:159–210.

Fellbaum, C. 1998. *WordNet - An Electronic Lexical Database*. Cambridge MA: MIT Press.

Gergen, K.; Greenberg, M.; and Willis, R., eds. 1980. *Social Exchange: Advances in Theory and Research*. New York: Plenum.

Girju, R. 2003. Automatic Detection of Causal Relations for Question Answering. In *The Association for Computational Linguistics (ACL 2003) Conference, Workshop on "Multi-lingual Summarization and Question Answering - Machine Learning and Beyond"*.

Haspelmath, M. 2007. Further remarks on reciprocal constructions. In Nedjalkov, V. P., ed., *Reciprocal Constructions*. 2087–2115.

Heim, I. 1991. Reciprocity and plurality. *Linguistic Inquiry* 22:63–101.

Hobbs, J. 2001. Causality. In *The Fifth Symposium of Formalisms of Commonsense Reasoning*.

Kingsbury, P.; Palmer, M.; and Marcus, M. 2002. Adding Semantic Annotation to the Penn Treebank. In *The Human Language Technology Conference*, 252–256.

Knobe, J. 2003. Intentional Action and Side Effects in Ordinary Language. *Analysis* 63:190–193.

König, E. 2005. Reciprocity in language: Cultural concepts and patterns of encoding. *Uhlenbeck Lecture* 23.

Li, X., and Roth, D. 2001. Exploring evidence for shallow parsing. In *The Annual Conference on Computational Natural Language Learning (CoNLL)*, 107–110.

Malle, B. F., and Bennett, R. E. 2002. People's praise and blame for intentions and actions: Implications of the folk concept of intentionality. *Institute of Cognitive and Decision Sciences, University of Oregon* 2(2).

Maslova, E., and Nedjalkov, V. 2005. Reciprocal constructions. In Haspelmath, M.; Dryer, M.; Gill, D.; and Comrie, B., eds., *The World Atlas of Language Structures*, 430–433. New York: Oxford University Press.

Paul, M.; Girju, R.; and Li, C. 2009. Mining the Web for Reciprocal Relationships. In *Conference on Computational Natural Language Learning (CoNLL-2009)*, 75–83.

Pearl, J. 2000. *Causality*. Cambridge University Press.

Sahlins, M., ed. 1972. *Stone Age Economics*. Chicago: Aldine-Atherton.

Shaver, K. G. 1985. *The Attribution Theory of Blame: Causality, Responsibility and Blameworthiness*. Springer-Verlag.

Wilson, T.; Wiebe, J.; and Hoffmann, P. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *The Human Language Technology Conference*.