

The Social Actions of Successful Groups

David B. Bracewell and Marc Tomlinson
Language Computer Corporation
Richardson, Texas 75080
Email:{david,marc}@languagecomputer.com

Abstract—The success of a group is determined by a number of factors. Some of these factors, such as task difficulty and availability of resources, are out of a group’s control and are constant amongst competing teams working on the same task. One factor which is not constant is the social dynamics of the group. The mix of behaviors by individuals in the group, as seen through social relations, social actions, and social roles are key in determining a group’s success. In this paper, we examine if it is possible to determine the success, or status, of a Wikipedia article through the social dynamics of the associated discussions. We capture social dynamics using four higher level social phenomena and fourteen lower level social acts, which we define from prevailing theories of group success. We examine discussions around English and Chinese Wikipedia articles and find that using social acts can increase prediction over standard network metric approaches by 35.0% for English to 90.3% and 12.3% for Chinese to 88.7%.

I. INTRODUCTION

The success of a group is dependent on a number of factors. While some of these factors can be out of the groups control (e.g. lack of physical resources, information, outside forces, etc.), other factors are under the control of the group, such as the level of criticism, group leadership, or level of communication. One of the most critical factors for group success is the social dynamics, or behaviors, of individuals within the group [1], [2], [3]. Social dynamics define the actions and roles of individuals and relations between individuals within the group. How group members interact and perceive others in the group and the group as a whole plays an important role in the productivity of the members [4], [5], [6].

One area of research that has been widely used for studying the social dynamics of groups is Social Network Analysis (SNA) [7], [8]. SNA-based approaches are able to capture how patterns of communications between individuals correlate with the efficiency of the group. However, SNA-based approaches do not capture a nuanced understanding of the content of those messages. The content of the communications between individuals within a group can provide insight into the social attitudes, goals, and actions of the group participants. Knowing the collective social cognition of the individuals can provide significant insight into the state of the group and their likelihood for success. We believe that a combination of SNA and a deeper analysis of the discourse making up the social network will work synergistically.

In this paper, we investigate how social dynamics can be used to determine the success of a group. In particular, we examine if the social dynamics of groups in English and Chinese speaking Wikipedia discussions can determine their

success. We measure success as the group’s ability to reach their goal, which for Wikipedia is to create a high quality article. Article quality is measured using the status associated with the article: *Normal*, *Good*, or *Featured*. Featured articles represent the highest quality of articles in terms of writing, journalistic standards, proper citations and professionalism. Good articles are ones that are not quite good enough to be featured, but are deemed decent. Finally, the Normal articles are non-featured non-good articles and represent the majority ($\approx 99.5\%$) of Wikipedia articles.

We perform sociolinguistic analysis of Wikipedia discussion threads to capture the social dynamics of the group in the thread. We examine how the presence of higher level social phenomena and lower level social acts affect the group’s success. The higher level social phenomena directly indicate the social dynamics taking place in the group. In order to determine which social phenomena are important for determining group success, we must first examine what makes a group successful.

There are two main theories presented in literature regarding group success (measured in how likely they are to accomplish their goals) in individualistic societies, such as is found in groups communicating in English on Wikipedia. The first states that non-adversarial, or collegial, groups are more likely to reach their goals [3]. The second states that there is an optimal amount of conflict for groups to be most productive [2]. This second hypothesis has been further refined to suggest that only certain types of conflict in moderate amounts actually make groups productive [9]. The type and amount of conflict is group and task dependent. We synthesize these theories into one general hypothesis for groups communicating in English (H_{EN}) as follows:

H_{EN} : A successful group is one which is mostly *collegial*, but has moderate amounts of certain types of *conflict*.

For groups communicating in Chinese, we found one central theory around group success in the literature; Groups who are non-adversarial and who have strong leadership are more likely to be successful. He (2005) draws similarity between groups and military units stating that for a group to be successful, they need to have little adversarial behavior and strong leadership [10]. Tao and Li state that groups in which individuals work together and are less adversarial can improve group performance [11]. They further state that weak leaders, e.g. ones with no recognition, cause a group to fail. We

summarize the works of He and Tao and Li into one general hypothesis for groups communicating in Chinese (H_{ZH}) as follows:

H_{ZH} : A successful group is one which is *collegial* and has strong *leadership*.

Using H_{EN} and H_{ZH} , we identify four social phenomena for identifying if a group will be successful: (1) *Collegiality*, (2) *Adversarial Behavior*, (3) *Leadership*, and (4) *Pursuit of Power*. Collegiality and Leadership are directly referenced in both hypotheses. Adversarial Behavior and Pursuit of Power are types of conflict with Adversarial Behavior representing relationship conflict and Pursuit of Power being task conflict.

To infer the higher level social phenomena, we use social acts. Social acts are captured using the language employed by individuals in the discussions. We have identified fourteen social acts, listed in section III, from the fields of psychology and organizational behavior. Social acts are captured using a combination of lexical, grammatical, and semantic cues as well as gappy n-gram models.

Finally, we examine the effectiveness of social acts alone and coupled with network metrics in automatically determining the success of a group in Wikipedia. Namely, we examine if the deeper pragmatic knowledge facilitated by the social acts can improve the classification accuracy of group success over network metrics alone. We experiment with regression and classification to determine the ranking and predictive power of the social acts.

II. RELATED WORK

As the popularity of Wikipedia has grown so has the amount of research around it. One area that has seen an influx in the past few years is on the effect collaboration has on the quality of the articles. Typically, collaboration has been examined using the edit history associated with each Wikipedia page. A common approach is to build a social network from the interaction in the edit history and to analyze the networks using various metrics. The computed metrics are then often used as features in a machine learning algorithm for prediction of quality.

Wang and Iwaihara (2011) propose a model using network structure of the editors to predict the quality of the articles [12]. The network structure is built from the interactions in the edit history. They use a number of network structural indicators, including the *restoreratio*. The *restoreratio* is used to determine if restores, the act of reverting a change in a Wikipedia page, are done between friendly or hostile group members. The metrics are used as features in a support vector machine in order to determine the quality of an article.

Other SNA based approaches have focused on certain aspects of behavior in the edit history. Hu et al. (2007) investigate assigning quality to Wikipedia articles using three models [13] that use information from edit history interactions of contributors. Their models take into account author authority, peer review behavior, and partial reviews.

Other areas of research have focused on specific types of social actions that are predictive of success. Kittur and Kraut

(2008) examine how coordination between editors affects article quality [14]. They categorize coordination as explicit and implicit. Explicit coordination is a planned effort among all editors whereas implicit coordination is when a subset of the editors plan and perform most of the work. They found that adding more editors to an article only increases quality when using an effective form of coordination and otherwise reduces the quality.

Closer to the research presented in this paper is the work by Nemoto et al. (2011) which examines how social capital impacts the efficiency of collaboration in Wikipedia [15]. They construct a social network using interactions in the associated talk pages. They examine featured vs. non-featured articles and measure the time promotion to featured article as their metric. They found that articles that were promoted faster had more cohesive and centralized groups with a larger number of members.

Also pertinent to the research presented in this paper is work done in the areas of social relationship extraction and social act analysis. Work in the area of social relationship extraction can be divided into several areas. The field of socio-linguistics boasts well-established studies of interpersonal relationships. For example, Eggins and Slade present a thorough linguistic analysis on causal conversations that covers topics such as humor, attitude, friendliness, and gossip [16]. This is accomplished through a comprehensive analysis of the dialogue at multiple levels.

One way in which discourse can be analyzed is through the use of dialogue acts. There are a number of schemes for coding dialogue acts, such as DAMSL [17] and VERBMOBIL [18]. However, dialogue acts are not capable of capturing the nuances of the social intentions of the discourse participants. Instead of focusing solely on the discourse, we must also focus on the discourse participants and how their social aspirations and goals, as captured through social acts, affect the productivity of themselves and others around them. Capturing the social goals of individuals in a group can inform to the likelihood of the group's success.

Social acts focus on the social implicature of the statement and thus more directly relate to the social intentions and goals of individuals. Researchers have recently begun to construct and annotate social acts. Bender et al. (2011) create an annotated corpus of social acts relating to *authority claims* and *alignment moves* [19]. They report kappa values in the range of 0.13 to 0.63, which is comparable to the 0.15 to 0.76 kappa values for dialogue acts reported by Allen and Core [17].

III. MODELING SOCIAL DYNAMICS

Based on the predominate hypotheses around social dynamics and group success as discussed in the introduction, four high level social phenomena and fourteen social acts are used to identify actions, roles, and relationships in discussions.

A. Social Acts

Task, or goal, oriented dialogue has many possible social implications. These social implications can include expres-

TABLE I: The fourteen social acts used to identify Adversarial Behavior, Collegiality, Pursuit of Power, and Leadership.

Social Acts	Definition
Acknowledge Shared Values	Common beliefs, such as religious, political, ethnic, gender, or moral, that are shared by individuals.
Negate Shared Values	Common beliefs for which two individuals are in direct opposition over.
Mutually Supportive Behavior	Statements of support for another’s ideas/actions manifested through encouragement, intention to help, praising, etc.
Undermining	Hostile expressions meant to erode the support or weaken the stance of another’s goal.
Establish Solidarity	Explicit affiliation by one individual toward another by drawing attention to attributes or connections shared.
Offer Gratitude	A sincere expression of thanks by one individual toward another.
Disrespect	Inappropriate language, such as mocking, or derogatory statements.
Agreement	Explicit utterances of agreement, approval, or acceptance.
Disagreement	Explicit utterances of disagreement, disapproval, or rejection.
Challenge Credibility	Attempts to challenge or discredit another’s credibility.
Managerial Influence	Assertions of administrative sway or controlling the direction in the discussion.
Group Affordance	Shows of respect and esteem through honorifics and yielding to others in the group.
Establish Credibility	Attempts to increase standing within group through demonstrating experience.
Leadership Traits	Language, such as locus of control and certainty, associated with individuals who have a leadership role.

sions of desire to establish or reaffirm a bond between individuals or the desire to increase one’s power and sway with others. The types of social interaction that transpire within a group are predictive of the group’s ability to reach its goals. Thus, it is imperative to understand the social implications of individuals in a group discourse to predict the group’s likely success or failure. The social implications of a discourse can be determined by examining the social intentions, or actions, of the discourse participants.

We label the social intentions of an utterance, as social acts. Social acts can range from establishing mutual bonds to asserting dominance over another individual. They can be signaled with a variety of cue phrases as well as through a discourse participant’s observation or violation of social norms, or their expectation of a socially appropriate response.

We define a set of fourteen social acts, shown in Table I, that inform the four social phenomena (Adversarial Behavior, Collegiality, Pursuit of Power, and Leadership) that we wish to capture. These social acts are derived from work in psychology on power, status, and leadership [20], [21], [22], [23], [24], as well as on conflict and cooperation [25], [26], [2]. Detailed descriptions of these acts can be found in [27], [28].

In order to capture the social acts, we first must have a corpus with an ample amount of social act annotations. Our corpus for English has 21,067 English sentences with 8,149 (38.7%) of them having one or more social acts annotated. For Chinese, our corpus consists of 24,339 sentences with 11,537 (47.4%) of them having one or more social acts annotated. On average each English sentence is assigned 1.02 social acts and each Chinese sentence is assigned 1.77 social acts. Kappa values for English annotations range from 0.09 for Group Affordance to 0.52 for Establish Solidarity. For Chinese, the Kappa values range from 0.07 for Disagreement to 0.85 for Mutually Supportive Behavior. Figure 1 shows an excerpt of a discussion communicated in English with social acts tagged.

The annotated corpus of social acts is split into 80% for training and 20% for testing. A combination of hand-

Turn 1] Propose that this page be moved to East Timor Defence Force as this is the closest translation of Foras de Defesa de Timor Leste. I have worked in Timor Leste as a government advisor, including with FDTL, and have never heard anybody ever refer to the FDTL as Military of East Timor. P1

Turn 2] As I understand it, 'East Timor Defence Force' is considered outdated. **While it was commonly used when the force was established, almost all english-language publications now use 'F-FDTL'. 'Military of East Timor' is a generic name, and I agree that it's rarely used and not a great title.**[Agreement] I'd prefer 'Timor Leste Defence Force' as this seems to be the direct translation, but this would be inconsistent with the other Wikipedia articles on the country. **Should we be bold and move this article to 'Timor Leste Defence Force'?**[Establish Solidarity] P2

Turn 3] **I so totally agree with you.** [Agreement] **'Timor Leste Defence Force' is it.** [Agreement] **The only reason I did not propose that was the failure to change the country page from East Timor to Timor Leste, a decision that I feel was extremely discourteous of Wikipedia considering the government's specific request that it be referred to as Timor Leste.**[Establish Solidarity] If you have worked there you will know that everybody uses 'Timor Leste', even the ADF but the Australian DFAT uses East Timor although the more enlightened Kiwi embassy call it TL. I suggest we leave it for 48 hours to see if anyone has any strong feelings and then change it to ' Timor Leste Defence Force' with diverts from F-FDTL and FDTL P1

Turn 4] **I agree with that approach.** [Agreement] In the interests of consensus editing, I've posted a note at Talk:East Timor (in lieu of a Wikiproject on the country) to seek other editors' views. P2

...

Turn 8] **As no-one has raised any objections, I've just made the move.**[Mutually Supportive Behavior] P2

Turn 9] **Good move, well done!**[Mutually Supportive Behavior] P1

Fig. 1: An example discourse communicated in English with social acts labeled.

crafted lexical, grammatical, and semantic rules are written for both English and Chinese over the 80% training data. In addition, a gappy n-gram model, similar to that used for machine translation in [29], is trained. This combination of rules and gappy n-grams gives the best overall F-measure for classification of social acts. Our best models obtain a 50.4% F-Measure for English and a 52.1% F-Measure for Chinese.

Further detail on the annotation and classification process will be discussed in later papers.

B. Social Phenomena

The fourteen social acts are mapped into higher level social phenomena, which give an aggregated view of how individuals are participating within in the group. The literature on group success and social dynamics for groups communicating in English [2], [3] and groups communicating in Chinese [10], [11] suggest four social phenomena of importance: (1) *Collegiality*, (2) *Adversarial Behavior*, (3) *Leadership*, and (4) *Pursuit of Power*.

Collegiality is defined as cooperating with others in order to reach a common goal or ideal. Collegiality is directly referenced in both H_{EN} and H_{ZH} as being a positive indicator of groups success. In contrast, Adversarial Behavior is defined as direct opposition and disagreement with others, which may be hostile in nature. Adversarial Behavior can be seen as a form of relationship conflict [30].

The set of social acts mapping into Adversarial Behavior are indicative of an individual being contentious or hostile with others; The “Adversarial” social acts include:

Disagreement	Explicit statements showing rejection of other individuals’ statements.
Disrespect	Disrespected individuals often feel they have been unjustly treated due to the disrespectful action, causing a social imbalance between them and the perpetrator [31].
Undermining	A form of hostility meant to erode the support or weaken the stance of another individual.
Negate Shared Values	Signifies opposition over the goal and not belonging to a common group.

The set of social acts mapped into Collegiality are indicative of sharing a common bond or focus with other group members and a desire to work for a greater good; The “Collegial” social acts include:

Ack. Shared Values	Shared values suggest common group membership, which is indicative of a collegial relationship between the individuals.
Supportive Behavior	Supportive behavior, or cooperation, is correlated with both overall group performance and managerial ratings of group effectiveness [32].
Establish Solidarity	Language indicative of a desire for group solidarity encapsulates the establishment and maintenance of shared group membership.
Offer Gratitude	Even in the absence of any major differences within a group, the expression of an in-group bias and out-group bias [33] between individuals still takes place. Individuals within a group are more likely to possess positive feelings for another individual within the group and to rate him or her more highly than an individual outside of the group.
Agreement	Agreement can act as an affordance to an individual or as a means to establish solidarity between individuals.

Leadership and the lack of leadership is directly mentioned in the theories of group success for Chinese speaking groups [10], [11]. Here, we define leadership as the ability of an

individual to guide the group’s tasks and discussion toward a goal. When there is a lack of leadership or leadership is weak, individuals may try to pursue power. Pursuit of Power is the attempt by an individual to increase their status or sway in the group. Pursuit of Power often results in conflict as individuals within a group vie for a limited resource [30]. These conflicts can occur between individuals of the same status (colleagues) or individuals of different status (manager and subordinate).

The set of social acts mapping into Pursuit of Power are indicative of an individual trying to gain power in a group through either increasing their own status or lowering others’ status; The “Pursuit of Power” social acts include:

Establish Credibility	Establishing credibility reflects an attempt by an individual to demonstrate their credibility and fitness for leadership.
Challenge Credibility	One way individuals seek power is to lower the status of other group members. These challenges can be in demands to prove credibility and aggressive accusing questions.

The set of social acts mapping into Leadership are indicative of an individual who has personality traits associated with being task leader (e.g. locus of control and certainty), manages the group, and receives acknowledgement from the group; The “Leadership” social acts include:

Managerial Influence	Is used by individuals to signal that they are a leader by showing they have administrative sway or control over the group.
Leadership Traits	Language usage that is often associated with individuals in leadership.
Group Affordance	The affordance of power by group members to an individual is a sign of that individual’s power.

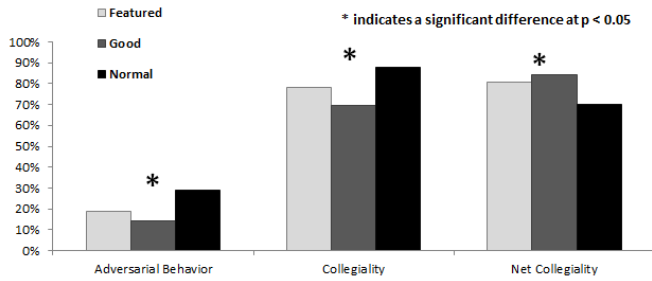
IV. DATA ANALYSIS

We collected 100 random discussions, containing 20 to 40 turns per discussion, from Wikipedia for each level of success (Normal, Good, and Featured). In total there were 300 discussions communicated in English and 300 discussions communicated in Chinese. Each discussion thread was analyzed for the fourteen social acts shown in Table I. Summary statistics for the number of turns, interactivity (percentage of turns that were replies), and the percentage of turns in which a social act was employed is shown in Table II.

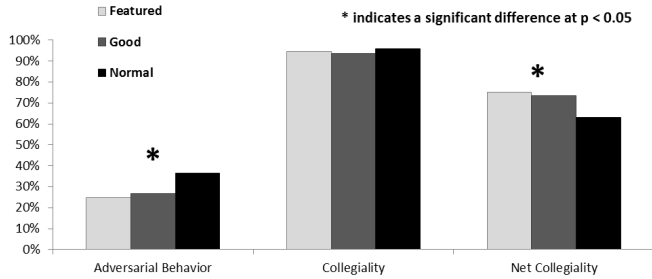
TABLE II: Summary statistics for the Wikipedia discussions threads by language and level of success.

	Success	Avg. turns	Interactivity	Pct. of turns with a social act
English	Normal	41	74%	97%
	Good	34	49%	41%
	Featured	32	51%	65%
Chinese	Normal	26	36%	84%
	Good	32	21%	59%
	Featured	23	14%	60%

As is seen in Table II, the amount of interactivity in discussions generally decreases as a higher level of success is



(a) Discussion threads communicated in English.



(b) Discussion threads communicated in Chinese.

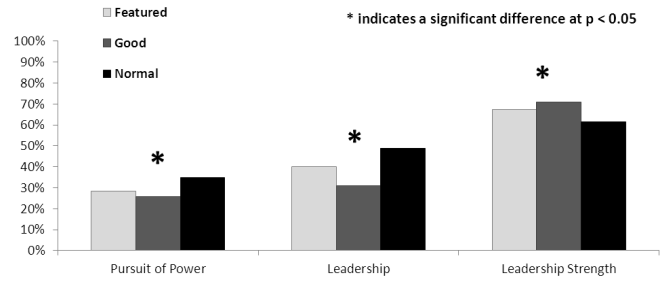
Fig. 2: The rates of Adversarial and Collegial social acts in discussion threads. (* indicates significant χ^2 at $p < 0.05$)

reached for both English and Chinese speaking groups. This is more evident in discussions communicated in Chinese. An examination of the data revealed that the lower interactivity in the Chinese speaking discussions was because a large number of the discussions for Good and Featured articles involved voting over proposed changes to the article. These voting discussions were merely “thumbs up” or “thumbs down” votes with little to no dialogue, which also explains the decrease in the percentage of turns with social acts.

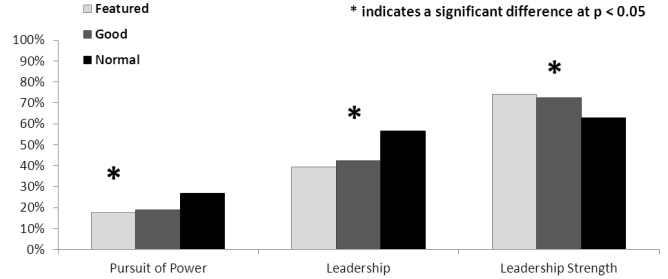
The percentage of turns with a social act for discussions communicated in English was less in discussions around Good articles compared to discussions around Normal and Featured articles. Analysis of the social phenomena will show if the increased usage of social acts aligns with the notion of moderate amounts of conflict as found in H_{EN} .

One commonality between H_{EN} and H_{ZH} is that they both reference the need for a successful group to be collegial. However, H_{EN} allows for moderate amounts of certain types of conflict. Figure 2 shows the rates of social acts (percentage of turns) indicating Collegiality and Adversarial Behavior. In addition, the Figure shows the “Net Collegiality” which is calculated as the percentage of turns exhibiting Collegiality that did not also exhibit Adversarial Behavior.

As seen in Figure 2, there are significant differences between the rates of Adversarial Behavior in discussions communicated in both languages and for Collegiality in discussions communicated in English. For discussions communicated in English, the rate of Collegiality is lowest in discussions for Good articles, but the Net Collegiality rate is the highest in



(a) Discussion threads communicated in English.



(b) Discussion threads communicated in Chinese.

Fig. 3: The rates of Leadership and Pursuit of Power social acts. (* indicates significant χ^2 at $p < 0.05$)

Good article. This means that the Collegiality expressed in discussions for Normal and Featured articles is often accompanied by some type of Adversarial Behavior. The rate of Adversarial Behavior is the highest in discussions for Normal articles, which follows H_{EN} . The discussions for Featured articles had the second highest rate of Adversarial Behavior. The Adversarial Behavior expressed in these discussions was mostly in the form of Disagreement.

For discussions communicated in Chinese there is no significant difference between the amounts of social acts indicating Collegiality between the three levels of success. However, the differences in Adversarial Behavior social acts is significant and steadily decreased as the level of success increased. In addition, higher levels of success also have a higher Net Collegiality, which corroborates H_{ZH} .

Another form of conflict, Pursuit of Power, is present when there is a lack of leadership or when the leadership is weak [30]. Figure 3 shows the rates of social acts (percentage of turns) indicating Pursuit of Power and Task Leader. In addition, the strength of leadership is shown in Figure 3. The Leadership strength is calculated as the rate of turns that exhibited a Leadership social act and did not exhibit a Pursuit of Power social act. Often times individuals pursuing power act like they are already in power. Thus, we interpret turns in which Pursuit of Power and Leadership social acts are exhibited as an intention to pursue power.

For discussions communicated in Chinese, the rates of social acts indicating Leadership and Pursuit of Power decreases as the level of success of increases. However, the strength

TABLE III: Classification results for modeling success.

		Social Acts			Network Metrics			Combined		
	Success	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
English	Normal	79.1%	72.0%	75.4%	68.1%	81.8%	74.3%	93.1%	88.2%	90.5%
	Good	88.5%	92.0%	90.2%	57.9%	73.0%	64.6%	87.8%	88.7%	88.2%
	Featured	64.9%	68.5%	66.7%	38.9%	21.0%	27.3%	90.9%	93.8%	92.3%
	micro-Avg.	77.9%	77.7%	77.7%	54.9%	58.5%	55.3%	90.4%	90.3%	90.3%
Chinese	Normal	52.5%	40.3%	45.6%	100.0%	97.4%	98.7%	88.7%	79.7%	84.0%
	Good	42.2%	35.4%	38.5%	70.7%	58.6%	64.1%	86.7%	86.7%	86.7%
	Featured	45.9%	61.4%	52.5%	65.5%	78.0%	71.2%	90.7%	97.0%	93.8%
	micro-Avg.	46.4%	46.2%	45.6%	77.1%	76.5%	76.4%	88.8%	88.8%	88.7%

of Leadership increases. The increase in Leadership strength corresponds to the the notion of successful groups communicating in Chinese having strong leadership as detailed in H_{ZH} . Normal articles had the highest rate of Leadership social acts at 56%. This leadership, however, was weaker, i.e. had less group recognition and occurred more with Pursuits of Power, than in discussions for Featured and Good articles. The Leadership in discussions for Featured articles was the strongest (more recognition by the group) with Leadership in discussions for Good articles only slightly weaker.

For English, the rate of Pursuit of Power and Leadership show similar trends as Adversarial Behavior and Collegiality (shown in Figure 2a) in that the lowest rate is in discussions for Good articles. The Leadership strength is the strongest in discussions for Good articles and slightly weaker in discussions for Featured articles. Pursuit of Power, as defined in this paper, can be seen as a task conflict. Task conflict in moderate amounts has been shown to correlate to higher quality articles in Wikipedia [34]. Furthermore, the increase in conflict aligns with H_{EN} .

We found that for Wikipedia discussions communicated English aligns with H_{EN} . Groups that reached some success, as defined as the Wikipedia article reaching good status, have very low amounts of Adversarial Behavior and Pursuit of Power. Teams who reach the pinnacle of success, Featured status, have some Adversarial Behavior and Pursuit of Power that is driven by conflict. Manual analysis of the data reveals that the conflict found in discussions for Featured articles is not personal in nature, but instead was around ensuring article quality. Wikipedia discussions communicated in Chinese aligns with H_{ZH} . The discussions for Normal articles have more social acts indicative of Adversarial Behavior and have weaker leaders. Discussions for Good and Featured articles have less Adversarial Behavior than Normal articles. Similarly, Leadership in discussions for Good and Featured articles is stronger, i.e. had more group recognition, than in Normal articles.

V. MODELING SUCCESS

The data analysis discussed in section IV revealed that the four social phenomena (Adversarial Behavior, Collegiality, Pursuit of Power, and Leadership) and the fourteen social acts can distinguish differing levels of success. This section looks

at whether or not the social acts can be used to build an accurate computational model of group success. In particular, we examine if the social phenomena as manifested through social acts in one discussion of moderate size (20 - 40 turns) can predict the level of success for the article. We use both regression and classification in order to judge the separability and predictability of the social acts.

A. Regression

Regression allows examining the power of the language uses to separate the three levels of success. The levels of success are assigned a numeric value corresponding to the inverse of their rank of success, i.e. Featured articles are assigned a 3, Good articles a 2, and Normal articles a 1. The values of the social acts are the amount that each was employed in its associated discussion, i.e. the sum of the usage.

Linear regression is performed using least squares estimation over the entire data set using for each language (300 discussions communicated in English and 300 discussions communicated in Chinese). The regression results in an $R^2 = 0.28$ for discussions communicated in English and an $R^2 = 0.21$ in discussions communicated in Chinese. The predictive accuracy of the model for English speaking discussions is 48% and 41% for Chinese speaking discussions (test on train). Both of the results are higher than the baseline of 33%. The mean squared error (MSE) is 0.47 for English speaking discussions and 0.75 for Chinese speaking discussions. The values of the MSE indicate that most errors are between consecutive level of success, e.g. a Normal article misclassified as Good.

The regression results (R^2 values) indicate that the social acts are capable of separating the levels of success. In addition, the regression results suggest that the social acts should be valuable features in a classification approach to determining the level of success of an article.

B. Classification

Based on the regressions results, showing that social acts can separate levels of success, we also experiment with using classification to determine the level of success for an article. We examine using the complete set of fourteen social acts, listed in section III as well as network metrics. In particular, we wish to determine if the deeper pragmatic knowledge captured by social acts can improve classification accuracy over using

only network metrics. Classification is performed using multi-class logistic regression¹. Testing is performed using leave-one-out cross-validation. Table III shows the results for each level of success as well as the micro-average for English and Chinese discussions.

First, we examine using a set of five network metrics and one discourse metric, which relate to the work discussed in section II, for determining group success. We model the discourse as a weighted undirected graph with authors as nodes and edges existing between two individuals if they interacted with one another. The weight of the edge is the number of times they interacted. We calculate the following five network metrics : (1) the number of authors (nodes), (2) the number of interactions (sum of edge weights), (3) the density of the network, (4) the clustering coefficient of the whole network, (5) the diameter of the network.

The number of authors (editors in Wikipedia), which is the number of nodes in the network, has been found to be a good indicator of article quality [14]. The number of interactions, or the sum of the edges weights in the network, is an indicator of group participation. Network density is a ratio of number of ties (edges) in the network to the total number possible. We use density as an indicator of group cohesiveness, where more dense networks have most individuals participating with one another. We use the clustering coefficient of the whole network [35], [36]. The clustering coefficient is a measure of likely two nodes (authors) are to be clustered together. Higher network-wide clustering coefficients correlate to small average distances between nodes. Another measure of distance is the diameter of the network, which measures the greatest distance between any two nodes. In addition to these network metrics we use the number of turns in the discussion as a discourse metric. As seen in Table III the results of using network metrics as the only feature results in a micro-average F-Measure of 55.3% for English and 76.4% for Chinese.

Next, we examine using only the social acts as features. The values for these features are the total number of times the social act was employed in the associated discussion. As can be seen in Table III, English has a micro-average F-Measure of 77.7% and Chinese has a micro-average F-Measure of 45.6%. In both cases, the classification with social acts as features outperforms the baseline (the base rate of the classes) of 33.3%. The classification of success for English is aided by the fact that discussions around Good articles have a lower percentage of social acts per turn, which acts as a natural boundary between Featured and Normal articles. In contrast, there is very little distinction between discussions for Good and Featured articles in Chinese, which can be seen in Figures 2b and 3b.

Comparing the network only and social act only results, we see that for English, using the network metrics as the sole features has a decrease in F-Measure of 22.4% over using only the social acts. This decrease suggests that the

network structures of the discussions communicated in English vary little across the levels of success. In contrast, using the network metrics alone increases the F-Measure by 30.8% over using only the social acts for Chinese. One possibility for this increase is the presence of voting discussions in the Chinese data. The percentage of voting discussions increases with the level of success with no voting actually taking place in our set of Normal articles. Characteristics of these voting discussions includes a low amount of interaction, which translates into a low number of edges in the network. This causes most of the network metrics to have a steady decrease as the level of success increases, i.e. it makes for an easy decision boundary.

Finally, we examine the combination of using social acts and network metrics together. We created a joint regression model using features from both approaches. The idea being that social acts and network metrics capture different subtleties in what makes a group successful and that the combination will work synergistically. As seen in Table III, combining the social acts with the network metrics increases the F-Measure of 35.0% in English and 12.3% in Chinese over network metrics alone. The results show that the deeper pragmatic knowledge about the intentions and goals that comes from the social acts enhances the performance of the network metrics. Conversely, the network metrics add further evidence for the level of success of an article that the social acts do not capture.

VI. CONCLUSION

We have examined how social dynamics can be used to predict group success. This was done by examining the effect that the social dynamics in Wikipedia discussions communicated in English and communicated in Chinese had in determining the success of the corresponding article. We started with hypotheses for what makes a group successful. From these hypotheses we defined four social phenomena (Adversarial Behavior, Collegiality, Leadership, and Pursuit of power), which describe the social dynamics. A set of fourteen social acts, which relay the social intentions of the speakers, were used to capture these social phenomena.

We showed that there is a significant difference in the rate of the four social phenomena in the discussions between the levels of success. Our analysis corroborated the hypothesis for English (H_{EN}) that while collegial groups will gain success, there is a need for certain types of conflict to be present for the group to achieve even greater success. Similarly, we saw that the dynamics in Wikipedia discussions communicated in Chinese also align with the hypothesis for Chinese (H_{ZH}), which states collegial groups with strong leadership are more successful.

We further went on to show that the deeper knowledge of the social dynamics gained from the social acts increases performance over using network metrics alone for classifying the success of a group. In particular, we showed an increase of 35.0% in F-Measure for English and a 12.3% increase for Chinese when combining social acts with network metrics.

¹We also examined using support vector machines, but found logistic regression gave the best results.

ACKNOWLEDGMENT

This research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), and through the U.S. Army Research Lab. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI or the U.S. Government.

REFERENCES

- [1] J. R. Hackman, *Leading Teams: Setting the Stage for Great Performances - The Five Keys to Successful Teams*. Harvard Business Review Press, 2004.
- [2] K. A. Jehn and E. A. Mannix, "The Dynamic Nature of Conflict: A Longitudinal Study of Intragroup Conflict and Group Performance," *Academy of Management Journal*, vol. 44, no. 2, p. 238, 2001.
- [3] J. W. Little, "Norms of Collegiality and Experimentation: Workplace Conditions of School Success," *American Educational Research Journal*, vol. 19, no. 3, pp. 325–340, Jan. 1982.
- [4] D. Cartwright, "The nature of group cohesiveness," in *In Cartwright D and Zander Z eds GROUP DYNAMICS New York Harper and Row 1968 pp 91110*. Tavistock Publications, 1968, pp. 91–109.
- [5] D. Forsyth, *Group Dynamics*. Wadsworth Publishing Company, 1999.
- [6] R. P. Vecchio, *Organizational Behavior: Core Concepts*. South-Western College Pub, 2005.
- [7] D. Liben-Nowell and J. M. Kleinberg, "The link prediction problem for social networks," *Proceedings of the twelfth international conference on Information and knowledge management CIKM 03*, vol. 58, no. November 2003, p. 556, 2003.
- [8] J. Leskovec, D. Huttenlocher, and J. M. Kleinberg, "Predicting Positive and Negative Links in Online Social Networks," *Proceedings of the 19th international conference on World wide web WWW 10*, p. 641, 2010.
- [9] C. K. W. De Dreu, "The virtue and vice of workplace conflict: food for (pessimistic) thought," *Journal of Organizational Behavior*, vol. 29, no. April 2007, pp. 5–18, 2008.
- [10] X. He, *Jiefangjun Jingsheng, Zhongguo*. Fangzhi Chubanshe, 2005.
- [11] Y. Tao and J. Li, "How to enhance the effectiveness of collaborative learning of the group ." *Distance Education Journal*, vol. 6, 2005.
- [12] S. Wang and M. Iwaihara, "Quality evaluation of wikipedia articles through edit history and editor groups," in *Web Technologies and Applications*, ser. Lecture Notes in Computer Science, X. Du, W. Fan, J. Wang, Z. Peng, and M. Sharaf, Eds. Springer Berlin / Heidelberg, 2011, vol. 6612, pp. 188–199.
- [13] M. Hu, E.-P. Lim, A. Sun, H. W. Lauw, and B.-Q. Vuong, "Measuring article quality in wikipedia: models and evaluation," in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, ser. CIKM '07. New York, NY, USA: ACM, 2007, pp. 243–252.
- [14] A. Kittur and R. E. Kraut, "Harnessing the wisdom of crowds in wikipedia: Quality through coordination," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ser. CSCW '08. New York, NY, USA: ACM, 2008, pp. 37–46.
- [15] K. Nemoto, P. Gloor, and R. Laubacher, "Social capital increases efficiency of collaboration among wikipedia editors," in *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, ser. HT '11. New York, NY, USA: ACM, 2011, pp. 231–240.
- [16] S. Eggins and D. Slade, *Analysing casual conversation*. Cassell, 1997.
- [17] J. Allen and M. Core, "Draft of DAMSL: Dialog Act Markup in Several Layers," pp. 1–32, 1997.
- [18] S. Jekat, R. Klein, E. Maier, I. Maleck, M. Mast, T. Berlin, J. J. Quantz, and J. J. Quantz, "Dialogue acts in verbmobil," Tech. Rep., 1995.
- [19] E. M. Bender, J. T. Morgan, M. Oxley, M. Zachry, B. Hutchinson, A. Marin, B. Zhang, and M. Ostendorf, "Annotating social acts: authority claims and alignment moves in wikipedia talk pages," in *Proceedings of the Workshop on Languages in Social Media*, ser. LSM '11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 48–57.
- [20] C. Anderson, O. P. John, D. Keltner, and A. M. Kring, "Who attains social status? Effects of personality and physical attractiveness in social groups," *Journal of Personality and Social Psychology*, vol. 81, no. 1, pp. 116–132, 2001.
- [21] J. R. P. French and B. Raven, "The Bases of Social Power," in *Studies in Social Power*, ser. Studies in social power, D. Cartwright, Ed. Institute for social research, 1959, vol. 35, ch. 9, pp. 150–167.
- [22] D. Keltner, G. A. Van Kleef, S. Chen, and M. W. Kraus, "A reciprocal influence model of social power: Emerging principles and lines of inquiry," *Advances in experimental social psychology*, vol. 40, pp. 151–192, 2008.
- [23] D. Owens and R. Sutton, "Status contests in meetings: Negotiating the informal order," *Groups at work: Theory and research*, vol. 14, pp. 299–316, 2001.
- [24] P. Smith and A. Galinsky, "The nonconscious nature of power: Cues and consequences," *Social and Personality Psychology Compass*, vol. 4, no. 10, pp. 918–938, 2010.
- [25] M. B. Brewer and W. Gardner, "Who is this "We"? Levels of collective identity and self representations," *Journal of Personality and Social Psychology*, vol. 71, no. 1, pp. 83–93, 1996.
- [26] M. Deutsch, "Cooperation and competition," *Conflict, Interdependence, and Justice*, pp. 23–40, 2011.
- [27] D. B. Bracewell, M. T. Tomlinson, M. Brunson, J. Plymale, J. Bracewell, and D. Boerger, "Annotation of Adversarial and Collegial Social Actions in Discourse," in *6th Linguistic Annotation Workshop*, no. July, 2012, pp. 184–192.
- [28] M. Tomlinson, D. Bracewell, M. Draper, Z. Almissour, Y. Shi, and J. Bensley, "Pursing power in arabic on-line discussion forums," *Proceedings of LREC2012*, pp. 1359–1364, 2012.
- [29] K. Gimpel and N. A. Smith, "Generative Models of Monolingual and Bilingual Gappy Patterns," *Proceedings of the Sixth Workshop on Statistical Machine Translation*, no. 2009, pp. 512–522, 2011.
- [30] K. M. Keller, "Power Conflict: Struggles for Intragroup Control and Dominance," Ph.D. dissertation, University of Maryland, College Park, 2009.
- [31] D. T. Miller, "Disrespect and the experience of injustice," *Annual Review of Psychology*, vol. 52, no. 1, pp. 527–553, 2001.
- [32] M. Champion, E. Papper, and G. Medsker, "Relations between work team characteristics and effectiveness: A replication and extension," *Personnel psychology*, vol. 49, no. 2, pp. 429–452, 1996.
- [33] M. B. Brewer. "In-group bias in the minimal intergroup situation: A cognitive-motivational analysis." *Psychological Bulletin*, vol. 86, no. 2, pp. 307–324, 1979.
- [34] O. Arazy, O. Nov, R. Patterson, and L. Yeo, "Information quality in wikipedia: The effects of group composition and task conflict," *J. Manage. Inf. Syst.*, vol. 27, no. 4, pp. 71–98, Apr. 2011.
- [35] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *PROC.NATL.ACAD.SCI.USA*, vol. 101, p. 3747, 2004. [Online]. Available: doi:10.1073/pnas.0400087101
- [36] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.