ABSTRACT
Hundreds of millions of people leave digital footprints, including textual posts and photos, on public on social media and social networking sites. Here we present our work on using these digital footprints—social multimedia content—to derive four types of basic personal traits for individuals. Moreover, we show how these basic traits can be used in combinations to help assess composite traits in two cases: trust modeling and resilience modeling. We share our preliminary results and discuss future research directions.

Index Terms—social multimedia, psycholinguistics, physiognomy, social network analysis, personality, fundamental needs, social genome, basic value, trust, resilience

1. INTRODUCTION
A number of human studies from Psychology to Behavioral Economics have shown that personal traits, such as personality and motivation, are highly correlated with a person’s behavior and performance in the real world, including occupational proficiency and economic decisions (e.g., [10][32]). It is expensive or impractical, however, to extract these traits using traditional tests or surveys (e.g., Big 5 personality tests). For example, a business simply could not ask millions of its customers to take such tests. Moreover, standard tests or surveys may not exist for deriving certain types of personal traits, such as fundamental needs [10].

On the other hand, with the rapid growth of social media and social networking activities, hundreds of millions of people have left their digital footprints in public. A number of research efforts have begun to utilize these digital footprints to acquire a deeper understanding of individuals, including inferring demographics information [27], political orientation [27], and personality [14].

Inspired by these works, we are leveraging multimedia footprints left on social media sites (e.g., Twitter and Pinterest) to derive individual traits. Unlike most prior efforts, however, which focus on predicting people’s traits based on their online social activities, e.g., number of posts and votes, ours work examines people’s language and image choices in their social postings. Specifically, our personal trait prediction is based on two areas of research: (1) psycho-linguistic analysis that correlates personal traits with their linguistic footprints (e.g., [33][37]), and (2) physiognomic analysis that infers personal traits from image footprints (e.g., [3]).

In the first area, we develop a lexicon-based approach with a set of lexicons that predict psycho-metric scores based on individual word choices in written language. In the second area, we use image analysis to extract physiognomic features related to personal traits (e.g., [30][36]). Furthermore, we focus on deriving deeper psychological characteristics, such as fundamental needs and basic human values, which have not been computationally derived before.

So far we have focused on automatically deriving four types of basic personal traits: personality, fundamental needs, basic values, and social genome. These four types of traits not only capture psychological characters that reflect individual differences, but they are directly linked to one’s behavior and performance in the real world. For example, fundamental needs are related to one’s economic decisions [10], while basic values are directly linked to one’s motives [32]. Furthermore, we “compose” the basic traits to predict one’s composite traits. For example, we predict an individual’s level of resilience under stress by a combination of personality, basic values, and social genome. Figure 1 shows the personal traits of an individual derived by our system.

In the rest of the paper, we first provide a brief overview of related work before describing our system, which automatically collects and analyzes the social multimedia data generated by an individual to infer personal traits. We then describe how we model each type of personal trait from linguistic footprints. Based on the basic personal traits, we show how to model two composite traits with additional

Figure 1. A screen shot of our system showing the derived personal traits for a specific individual.
data, including social activities and images. One is to model the psychological characters of a trusteer—source of messages—that helps instill the trust in a truster—recipient of the message. The other is to model resilience, how well an individual can handle stresses and cope with adversary situations. Finally, we present our preliminary results and discuss current limitations and future directions.

2. RELATED WORK

Our work is related to several areas of research in modeling and analyzing people’s traits from various data sources. First, there is a rich body of literature on using social media and social network structures to derive behavior and traits (e.g., tie strengths [12] and personality [14]). Most existing works utilize social media features, such as user-input profile and the number of “likes”, our analysis uses content-based features, such as the types of words or images users choose in expressing themselves. We also aim to develop novel computational methodologies for acquiring deeper psychological portraits of individuals, e.g., predicting motivation and resilience.

Our approaches for extracting traits from textual input is built on existing work in psycholinguistic analytics (e.g., [33][37]). While we leverage existing psycholinguistic findings, e.g., the LIWC dictionary developed in [33], we also significantly extend prior work by constructing our own psycholinguistic dictionaries to derive traits which have never been computationally modeled before, such as fundamental needs and basic human values.

Our work on extracting personal traits from images is based on a rich body of work on physiognomy, the assessment of one’s personality from the appearance of their face [30][36]. Researchers have associated physical facial features (e.g., soft jaw line and complexion) to psychological traits (e.g., trustworthiness [13] and success [30]). However, existing associations are qualitative. Our work attempts to correlate facial features that are measurable from images analysis with psychological traits to automatically infer personal traits.

We ground our personal trait analysis in various findings in human studies on understanding and modeling people’s intrinsic traits (e.g., [10][21][32]). For example, our needs model is based on Psychology [19] and Marketing [10] that people’s needs are universal and hierarchical. Similarly, our work in “composing” basic personal traits to infer one’s composite traits is also grounded in human studies (e.g., modeling a trustee’s psychological characters [21] and a person’s resilience [16]). While we leverage the findings in human studies to identify personal traits to model, we build computational models to automatically infer these traits from an individual’s social multimedia footprints.

3. OVERVIEW OF OUR SYSTEM

To derive an individual’s personal traits from social multimedia, our system consists of four key components (Figure 2). The heart of the system is a data store for holding profile information about individuals, caching data for analysis, storing the analytical results, and providing data to the web-based UI. The data collection component gathers user-generated content from various social media sources (e.g., Twitter), processes it, and stores it for subsequent analysis. The data analysis component consists of a set of personal trait analytic appliances, each of which is used to derive a particular type of personal trait from the collected data. Currently, we have created four analytic appliances: personality, fundamental needs, human basic values, and social genome. These traits can be explored using our web-based user interface with interactive visualization.

3.1 Data Collection

Currently, we have collected and analyzed data primarily from Twitter, and also from IBM Connections (an enterprise collaboration platform which includes forums and blogs). We have collected and analyzed about 175 million tweets from the public profiles of about 370,000 Twitter users, and also about 2700 blogs and posts from 600 discussion groups from Connections. For Twitter, we have been collecting data using their free API, which provides comprehensive access to individual data, though in a rate-limited fashion.

For comparison purposes, we selected individuals who are the followers of various businesses, including retail, hospitality and certain government twitter accounts. Since not every user on social media is a human being (e.g., news sites or spammers), we have used a set of filtering criteria to exclude non-humans, e.g., avoiding accounts where all the posts contain URLs (as occur with spammers). Different types of analysis may demand additional filtering. For example, personality analysis requires a minimum amount of user-generated text. We thus filter out those who only retweet others’ posts.

3.2 Interactive Visual Analytic UI

To help a user examine the derived personal traits (his own or others), we have developed a web-based user interface supporting interactive visualization of derived traits. Not only can a user interactively drill down on the specific as-
4. DERIVING BASIC PERSONAL TRAITS

We have used a mix of empirical and computational approaches to model and derive each personal trait from social multimedia data. In general, we use a lexicon-based psycholinguistic analysis to process linguistic footprints and are experimenting with various image analysis techniques to derive the features that help reinforce the derived traits from the linguistic analysis. Next, we describe how we derive each type of personal traits.

4.1 Personality

Personality consists of a set of psychological characteristics that reflects individual differences. We chose to model personality for two main reasons. First, personality is consistent and enduring. Second, personality is directly linked to many aspects of life [14]. Moreover, psycholinguistic literature has already demonstrated the ability to infer personality from an individual’s linguistic footprints [33][37], including inferring personality from social media [14]. We thus have used the LIWC dictionary [33] and its extension [37] to infer the Big 5 personality traits and their sub-facets. The detailed calculation is described in our recent work [18].

4.2 Fundamental Needs

Research in Psychology [19] and Marketing [10] has identified a set of human fundamental needs, which forms a hierarchy. For example, a person’s needs ranges from basic physiological needs such as breathing, to higher-level safety needs (e.g., well-being), and to interpersonal needs of feeling belongingness. However, there is no established computational model that can infer fundamental needs from an individual’s digital footprints. We thus have employed a hybrid empirical and computational approach to model one’s needs from a person’s linguistic footprints (e.g., tweets and blogs).

First, we designed and conducted large-scale, psycho-metric studies to collect ground-truth training data. In particular, we model fundamental needs along the 12 dimensions described in [10]. For each need dimension, we designed an item-based survey that collects a person’s psycho-metric scores describing needs along that dimension. In the survey, we also collected participant-generated text describing a specific need. We collected the survey results from over 3000 participants recruited on Amazon Mechanical Turk. Second, based on the collected data, we then analyzed how the set of words extracted from participant-generated text are positively or negatively correlated with each dimension. For example, for the need dimension “ideas”, we found that words such as “accomplish” and “perfect”, are positively correlated with the dimension, while words like “bad” and “fix” are negatively correlated with the dimension. As a result, we constructed a dictionary characterizing the description of each needs dimension. Using the results, we then built a statistical model that can predict needs scores along the 12 dimensions based on one’s use of relevant words in textual footprints (e.g., tweets).

4.3 Human Basic Values

Human basic values describe motivational goals (varying in importance) that people strive to attain [32]. For example, some value “power” (social status and prestige), while others value “tradition”. We chose to model basic values to understand people’s motivations and corresponding incentives needed to incite them to act. For example, if one cherishes “power”, he will be more motivated to act if the actions will result in social esteem. Currently, we model ten distinct basic core values that are recognized across the world [32]. Similar to our method described above for inferring one’s needs through their word uses, we used a psychometric survey with 800 users to collect their value scores along with their posts. We then correlate each value dimension with the 68 LIWC word categories (e.g., family, positive feelings, and self-reference). The results then enable us to predict one’s scores on each value dimension based on their LIWC word uses in their linguistic footprints [6].

4.4 Social Genome

To better understand the make up of one’s social circle and gauge one’s influence within the circle as well as the support one expects from the circle, we study one’s social genome from two aspects. First, we model the tie strengths between a person and everyone in her social network since stronger ties signify greater influence and support [12][24]. Currently, we compute the tie strength between two people based on the frequency, recency, intensity, and intimacy of their interactions on Twitter, similar to the method used in [12]. Second, we model the intrinsic similarity between a person and everyone in her social network, since more similar people induce greater mutual trust and influence [20]. Currently, our similarity measuring takes into account the
three types of intrinsic traits described above: personality, fundamental needs, and basic values.

5. DERIVING COMPOSITE PERSONAL TRAITS
One of our ultimate goals for modeling basic personal traits is to use such traits and their combinations to model a person’s composite trait in a specific situation. Here we use two concrete use cases to demonstrate how we model a composite trait based on several basic intrinsic traits: (1) trust modeling that studies a person’s characteristics that help instill trust in others in socio-digital influence campaigns; and (2) resilience modeling that predicts one’s ability to cope with stress and adversity situations.

5.1 Trust Modeling
There are many studies on examining the credibility of social media messages [5] and more recent studies on investigating how people perceive the credibility of social media messages in various conditions [22][35]. There is little work on understanding, let alone using, the psychological characteristics of a trustee (source of the message) to gauge the credibility of social media messages. On the other hand, research in Psychology has shown that the psycho-characteristics of a source are important trust constructs [21]. Based on this work, we thus model three identified key psychological characters of a trustee: competence, benevolence, and integrity.

5.1.1 Competence
Competence measures a trustee’s ability to offer what a trustor desires. Our current competence modeling focuses on the person’s level of expertise. In particular, we use linguistic footprints (e.g., posts) to infer the person’s expertise on a particular subject [25]. Since such expertise is often context and domain dependent, we are also estimating competence by an individual’s overall confidence based on their image footprints [28]. Our work is to classify visual appearance of the trustee into domain-specific categories that induce different levels of trust. For example, a survey of 400 respondents indicated they were more willing (76.3%, P<0.0001) to share their problems with a physician who is dressed in professional attire with a white coat than a physician in surgical scrubs (10.2%), business (8.8%) or casual dress (4.7%) [29].

5.1.2 Benevolence
Benevolence measures how a trustee is motivated to act in the trustor’s interests. We compute benevolence from three aspects: basic values, personality, and responsiveness. First, we adopt the benevolence measure computed as one of the ten basic values. Second, we use the altruism measure, a facet of Agreeableness in our Big 5 personality model. Third, we compute responsiveness based on one’s social media activities, as it reflects one’s benevolence [21]. Currently, we measure response time—how quickly a person normally responds to a post directed to him/her on social media. We also compute proactiveness—how active a per-

Figure 4. Digitally-manipulated images of Gordon Brown showing (left to right) the original image, trustworthy features and untrustworthy features. (Source: The Telegraph 17 March 2013).

5.2 Resilience Modeling
Resilience refers to the ability to cope with painful events and unpleasant emotions in the real world. Quantitatively measuring one’s resilience level can then be used to help build resilience [16]. Although previous studies have identified multiple personal traits that contribute to resilience [16], our current work focuses on measuring two of them based on social multimedia data: social resilience [24] and emotional resilience [23].

5.2.1 Social Resilience
Social resilience examines an individual’s social relationships to others to gauges how these relations provide care and support for an individual [16][24]. In particular, two quality measures of one’s social network, structure and function, model the healthiness of the social network, which in turn signals a person’s ability to cope with stress and adversity. The structure of a social network measures the size and connection patterns (e.g., connectedness and frequency) in the network [24]. To characterize a structure, we leverage the concept of centrality to measure the number of direct
connections and stability of the network [8]. In contrast, the function of a social network measures the extent for one to receive physical or psychological support [24]. Computationally, it is estimated by the closeness of each link in the network. The closeness is computed by two parts: tie strength measured in one’s social genome, and the closeness score as part of the fundamental needs metric. Using weighted-degree-centrality [8], we can then combine both the structural (centrality) and functional (weight) measures. In case of sparse linguistic footprints, we also use images to model both structure and function of a social network. Specifically, we use an individual’s appearance or a lack of appearance in social settings (e.g., family and group photos) to measure tie strength. We also leverage a series of timestamped photos to model the evolution of relationships to others. The inferred closeness and stability from images can then be incorporated with the results from the text mentioned earlier to derive a more robust estimate of social resilience.

5.2.2 Emotional Style
In additional to social resilience, emotional resilience, measuring how fast an individual can recover from negative emotions, is another key contributor to one’s resilience [23]. Since emotional resilience has a strong correlation with emotional stability [26], we infer an individual’s emotional stability from both text and images. Currently, we use the LIWC categories related to emotion (e.g., anxiety and cheerfulness) and additional lexicons (ANEW) [4] to infer emotional state from text (e.g., tweets) during a particular time frame. We then build a temporal model that detects emotional state changes over time. In this model, we measure the frequency and degree of the changes to estimate one’s emotional stability.

In addition to inferring emotional stability from text, we leverage images as another predictor. We use off-the-shelf emotion classification of face images (e.g., [17]) to infer one’s emotional state. We then use the temporal model described above to infer one’s emotional state changes over time and estimate the frequency and degree of the changes.

6. PRELIMINARY RESULTS AND DISCUSSION
Since it is difficult and time consuming to validate our analysis results when there is a lack of ground truth (e.g., personality and actual tie strengths), here we report on N-fold cross validation results for two personal traits that we have ground truth: fundamental needs and basic values.

We ran 10-fold cross validation for the needs analysis against the collected data from our psycho-metric studies. Figure 5(a) shows the mean absolute errors of the model predictions from our. In addition, the R² values of the prediction for different needs were within the range of 0.1 to 0.2, similar to other psychometric prediction results such as personality from written text [14]. Figure 5(b) shows the correlations between predicted and the observed needs scores. The correlations are weak for certain needs (e.g., >0.5 in closeness), but are very poor for others (e.g., 0 for practicality). One of our hypotheses is that cultural factors might have affected the willingness for people to express certain needs in writing. For example, writing about mundane practical needs may not be common for the US population whom we surveyed. Overall, it is feasible to predict an individual’s certain needs from their writings. With more training data and feature engineering, it is foreseeable that better models can be built.

Our 10-fold cross-validation for predicting basic values from text showed similar results. The R² values for different dimensions were within the range of 0.14 to 0.18. The correlations between predicted and the observed value scores range from 0.35 to 0.41. More detailed analysis is in [6].

While we are collecting ground-truth data to validate two more basic traits (personality and social genome), we have run experiments to measure how much data is required to obtain stable results. Our analysis indicates that personality analysis of individuals is generally stable given more than 50-100 tweets. We analyzed people with 400-4000 tweets, computing their personality using their entire corpus, and compared that with varied sizes of tweet subsets. With only 50 tweets, the average error for personality traits was less than 10 percentile for more than 90% of the population.

We are also designing experiments to collect the ground truth of composite traits, which is much more tricky. Moreover, we are also investigating the correlations between different traits, as some of the psychological characteristics do overlap, e.g., one’s friendliness measured in personality and benevolence in basic values.

7. CONCLUSIONS
Leveraging social multimedia footprints, we have developed a system that automatically infers an individual’s personal
traits. In particular, we developed lexicon-based psycholinguistic analyses to derive an individual’s four basic personal traits from text: personality, fundamental needs, basic values, and social genome. Moreover, we combined the basic personal traits with physiognomic analyses of images to model two composite traits: psychological characters of a trustee and psychological resilience of an individual. Our preliminary results demonstrate the promise of our work.

REFERENCES

[37] Yarkoni, Tal. 2010. Personality in 100,000 words: A large-scale analysis of personality and word usage among bloggers. Journal of Research in Personality.