Ego-Centric Graphlets for Personality and Affective States Recognition

Stefano Teso*, Jacopo Staiano*, Bruno Lepri†‡, Andrea Passerini* and Fabio Pianesi†
*Department of Information Engineering and Computer Science
University of Trento, Italy
{teso,staiano,passerini}@disi.unitn.it
†Fondazione Bruno Kessler, Trento, Italy
{lepri,pianesi}@fbk.eu
‡Massachussets Institute of Technology, Boston, USA

***Abstract*—Do we tend to perceive ourselves more creative when surrounded by creative people? Or rather the opposite holds? Such information is very valuable to understand how to optimize work processes and boost people's productivity *along with* their happiness and satisfaction. Exploiting real-life data, collected over a period of six weeks in a research institution by means of wearable sensors, in this work we provide insights on human behavior dynamics in the workplace. We explore the use of graphlets, i.e. small induced subgraphs of a network, to encode the local structure of the interaction network of a subject, enriched with affective and personality states of his/her interaction partners. Our analysis shows that graphlets of increasing complexity, encoding non-trivial interaction patterns, are beneficial to affective and personality states recognition performance. We also find that different sensory channels, measuring proximity/co-location or face-to-face interactions, have different predictive power for distinct states.**

I. Introduction

Affect and personality permeate people's daily and working life and also the interdependent relationships they usually hold with bosses, colleagues, and subordinates. Several studies showed the relationships between personality and e.g. job performance [1] and job satisfaction [2]. At the same time, an affective revolution has taken place, in which academics and managers have begun to appreciate how an organizational approach that integrates employee affect provides a more complete perspective [3]. Previous studies outlined effects of affect on performance [4], decision making [5], and prosocial behavior [6].

Usually, we can think of affect and emotions both as *states* or *traits*. These two levels differ in terms of the extent to which they are deeply characteristic of the individual, and therefore the extent to which they are mutable or immutable. At the same time, traditionally scientific psychology has developed a view of personality as a higher-level abstraction encompassing traits, sets of stable dispositions towards action, belief and attitude formation. The problem with this approach to personality is that it assumes a direct and stable relationship between being extravert and acting extravertedly (e.g., speaking loudly, being talkative, etc.). Extraverts, on the contrary, can often be silent and reflexive and not talkative at all, while introverts can at time exhibit extraverted behaviors. While personality studies have often dismissed these fluctuations of actual behavior as statistical noise, it has been suggested by Fleeson [7], [8] that they can give a valuable contribution to personality prediction and to the understanding of the personality/behavior relationship. The social psychology literature has recently coined the term *personality states* to refer to concrete behaviors that can be described as having a similar content to the corresponding personality traits. In other words, a personality state describes a specific behavioral episode wherein a person behaves more or less introvertly/extravertly, more or less neurotically, etc.

In this paper, we investigate the influence played by specific situational factors, the face-to-face interactions and the proximity interactions with alters, over the ego's expression of a particular affective/emotional state or a specific personality state in a work environment. In particular, how the details and the complexity of the social network structure of the interacting alters can play a significant role in predicting the affective and personality states of the ego. To this end, we represent people's interactions as *graphlets*, induced subgraphs representing specific patterns of interaction, and design classification experiments with the target of predicting the subjects' self-reported personality and affective states. We investigate graphlets centered on the reference node (the *ego*), embedding information on the state of the alters and their interactions in order to recognize the affective/personality state of the ego. We explore how interaction patterns, encoded as graphlets, gathered from two distinct sensory channels, Bluetooth (BT) and infrared (IR), affect recognition of personality and affective states. Several studies in social psychology have revealed links between positive affect and social activity. Recently, Hatfield et al. coined the term *emotional contagion* [9] to describe the process by which people "catch" emotions from each other. Positive and negative moods also spread during long periods [10] and over workplace interactions [11]. Inspired by the susceptible-infected-susceptible (SIS) disease model, Hill et al. proposed a mathematical model for the contagion of long-lasting emotional states in a self-reported social network [12]. In social and ubiquitous computing, researchers have explored the associations between mood and social interactions captured by mobile phones [13]. These studies assume that for detecting or predicting if an individual is in a positive or negative affect state it is enough to look at the number of individuals with whom he or she is in contact, and possibly at their state. Instead, we investigate how the *structure* of the interaction network can play a significant role in predicting the affective and personality states of the ego. Regarding personality, researchers have started exploring the wealth of behavioral data made available by cameras and microphones in the environment [14], [15], smartphones [16], [17], wearable sensors [18] in order to automatically classify personality traits. However, the general approach of all these previous works is to isolate promising correlates of the targeted traits for classification or regression. All these works adopted the so-called person-perspective on personality and target personality traits prediction or classification and not personality states prediction or classification.

II. Sociometric Badges Corpus

For this study we exploited the SocioMetric Badges Corpus [19], a multimodal corpus designed to capture the psychological and situational aspects of the daily lives of employees in an organizational structure. The data were collected in a research institute for six weeks on a sample of 54 subjects during their working hours. Males predominated (90.8%) while the average age was 36.83 ± 8.61 years. The data were collected using wearable sensors called Sociometric Badges. These sensors are equipped with accelerometers, audio, Bluetooth and Infrared to respectively capture: body movements, prosodic speech features, co-location with other individuals and face-to-face interactions [20]. SocioMetric Badges have been used in several studies to capture face-to-face communication patterns, relationships among individuals, collective behavior and performance outcomes, such as productivity and job satisfaction [21]. An Experience Sampling Methodology (ESM) was employed to collect information about transient psychological states (personality and affectivity states). Participants completed a short Internet-based survey three times a day (at 11:00 AM, 2:00 PM, and 5:00 PM), automatically administered via email. Users were granted 2.5 hours to fill them. Participants were asked to confirm their availability during the 30 minutes before starting the questionnaire; only if confirmed, their responses would be included in the database. The questions in the surveys referred to affectivity and personality states experienced over the 30 minutes prior to the survey. The ten-item personality inventory TIPI [22] was used to assess personality states. A 7-point scale ranging from 1=*Strongly Disagree* to 7=*Strongly Agree* was used for responses. The scores for each state were calculated by summing the raw scores of the two corresponding items. Similarly, respondents were asked to report on a scale from 1 to 5 (1=*Very Slightly Or Not At All* and 5=*Extremely*) to what extent they experienced High Positive Affect (HPA) and/or High Negative Affect (HNA). Table I reports basic dataset statistics on the target behavioral states[1]. In this work, we exploit information from the infrared (IR) and Bluetooth (BT) sensors.

TABLE I. Mean and median daily values for personality and affective states in the dataset used.

state	mean	median
Extroversion	4.07	4.0
Agreeableness	5.13	5.5
Conscentiousness	5.53	6.0
Emotional Stability	5.54	6.0
Openness/Creativity	4.50	4.5
High Positive Affection	3.12	3.3
High Negative Affection	1.42	1.3

[1]These values have been obtained directly from the authors as errata corrige to the related publication.

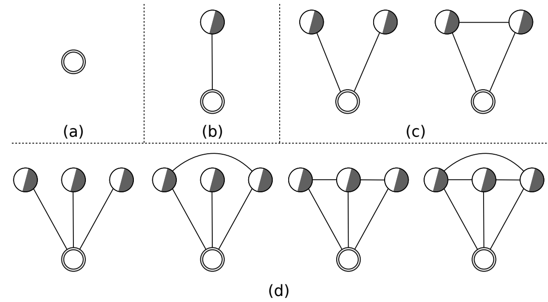

Fig. 1. Graphlet configurations used. Bottom nodes (double circled) represent the reference subject (the ego), while top nodes represent alters and their binary state.

III. Graphlet-based Approach

We define a binary classification task for each subject and each personality and affective state. This is done by mapping the state of a given subject at a certain deadline from $\{1, \ldots, 7\}$ to $\{0, 1\}$ using its median value for the subject as a threshold. Therefore, negative labels represent cases where the subject was found below his/her median. One of the main contributions of this paper lies in the encoding of the subjects' interactions as *graphlets*, defined as induced subgraphs of a larger network, providing a succinct representation of social structure. In the Bioinformatics and Computational Biology domains, graphlets have been introduced for the study of large biological networks, for e.g. network alignment [23]. Recently, graphlet analysis has been applied to Facebook messaging and historical crime data [24]. We investigate their effectiveness in the context of a human interaction network, for the prediction of behavioral determinants such as personality and affective states. Starting from the network of interactions between subjects, we extract for each subject the graphlets representing his/her local interactions. In this work, we consider all possible graphlets up to 4 nodes, as shown in Figure 1, where the double circle represents the reference subject and his/her interacting partners can have multiple patterns of reciprocal interaction. Furthermore, the graphlets embed information on the current (binary) state of the alters (but not of the reference subject whose state is to be predicted), in order to account for possible influence and propagation effects. For each deadline, we extract graphlet-based features from sensory data gathered over the previous 3 hours. We discretize each 3-hour window in 15-minutes slices in order to represent the evolution of the interaction patterns over time, taking into consideration the neighbours' states in order to account for situational influence effects. To do so, we count occurrences of graphlet configurations and build a histogram; then, we average the histograms obtained for each slice and obtain a feature vector representative of the 3-hours window under analysis. Finally, we use the latter to predict the ego's state at the deadline under analysis. In our setup, two kinds of missing data are possible: i) missing labels (i.e. surveys not filled by the subjects); ii) missing interactions, in which case the interaction graphs will be empty. In both cases we exclude the deadline under analysis from the training and testing stages.

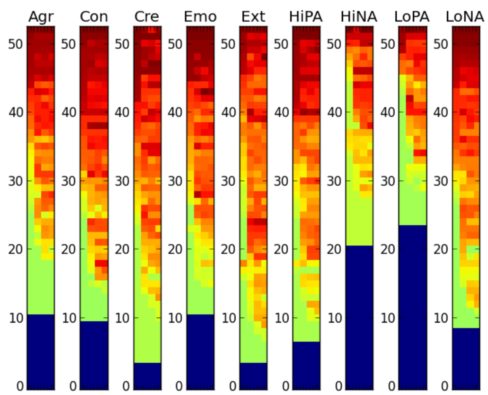

Fig. 2. Performance obtained using infrared sensory data.

IV. Experimental Setup

To understand the influence of alters on a given subject, we predict its state based on features derived from the labeled graphlets.

We build a linear Support Vector Machine (SVM) model [25] for each agent and each target state, evaluating its performance in a leave-1-week-out cross-validation procedure. We employed the LibLinear [26] library with ℓ_1 regularization, which tends to produce sparse models (with few non-zero weights). The learned models prioritize informative features, leading to robust handling of noise, and are simpler to interpret. To avoid any bias in the interpretation of the results, we discard all agent/target state pairs for which one class (i.e. positive or negative) covers at least 75% of the instances. This occurs when subjects exhibit very little variance on the labels (see Low PA - High NA in Figure 2), and thus many instances fall on the median value itself. We build models of increasing complexity by considering graphlet-based features made of up to one, two and three alters respectively, and evaluate the performance of each model in order to assess the predictive power of different levels of interaction. We compare these models against each other and against a majority classifier (i.e. a classifier that always predicts the class with more instances in the training set), which we use as a baseline.

V. Experimental Results

Figure 2 shows the performances obtained using features extracted from IR sensors (those for BT are similar). Each row represents a subject in our dataset. Each bar represents a target state. The left-most columns within each target display the performance of our baseline (a majority classifier), while subsequent ones represent those of the SVM models built using graphlets of increasing complexity. Each pixel represents the average f_1 measure [2] over all folds for a specific subject/target pair. The values range from low (light green) to high (red). Blue values represent cases with missing labels or highly unbalanced classes, which we ignore in the following analysis. For the sake of readability, we cluster the subjects (rows) by performance, using k-means with $k = 3$, and plot them from worst (bottom) to best (top). The plot shows that the graphlets are very predictive in the middle two clusters, where the performances tend to transition from bad (green) to good (red). This trend is not as clear on the high-performance cluster (top). The latter represents cases with an unbalanced, overly-positive data component (yet no more than 75% of the total), where the agents show little state variance and interactions can not be useful. To better understand the above trends we compute Win/Loss matrices for each target state and cluster: for each pair of feature-sets and for all subjects in a cluster we count the number of times a feature-set outperforms the others. Table II reports the resulting matrices for all target states and clusters for IR (results for BT are similar). Intuitively, positive values below the diagonal imply that the more complex feature-sets are informative and bring performance benefits; positive values above the diagonal have the opposite meaning. Analysis of the matrices shows that using more complex graphlets constantly improves the performances in the low and middle clusters. The behavior on the high-performance cluster is less clear, due to the aforementioned unbalance in the agent states distribution. We then compare the results of different target states for the two clusters (low- and middle-) where performances improve. For each matrix we compute the percentage of cases in which higher degree graphlets outperform (underperform) lower-degree ones, using the normalized sum of all elements above (below) the diagonal. Table III lists the difference between the two values, i.e., the relative percentage of cases for which more complex graphlets are beneficial. The results confirm that graphlet-based features are predictive of personality and affective states: all values are positive and well above the mean of a random classifier (i.e. 0). Finally, we sort such values for IR and BT, and compute the Spearman rank correlation between the two lists. The correlation coefficient is found to be -0.4 (p-value$= 0.28$), and indicates that the two channels can be effectively exploited for different target states, and support the intuition that the two channels capture different behavioral manifestations: BT captures proximity in a broad-cast manner (i.e. many-to-one), while IR is restricted to face-to-face (one-to-one) interactions.

TABLE III. Performance improvement factors, in ascending order. Value range is $[-1, 1]$.

IR		BT	
Target	Improvement	Target	Improvement
High PA	0.51	Creativity	0.44
Low NA	0.55	Conscientiousness	0.55
Extraversion	0.58	Emotional Stability	0.62
Emotional Stability	0.59	High PA	0.63
Creativity	0.6	Agreeableness	0.67
Agreeableness	0.63	Low PA	0.67
Low PA	0.69	Extraversion	0.69
High NA	0.70	Low NA	0.71
Conscientiousness	0.76	High NA	0.81

VI. Conclusion

In this paper we investigate new perspectives on affect and personality states recognition, studying in particular the influence on the ego's state of face-to-face and proximity interactions with alters. To this end, we propose a graphlet representation of the ego-network, computed using two distinct sensory channels (Bluetooth and infrared), to predict the ego's state. The advantage of graphlets over other representations is that they capture not only the number of interactions, but also their structure at different levels of complexity. Our results demonstrate that the graphlet-based representation consistently

[2] The f_1 measure is defined as the harmonic mean of precision and recall.

TABLE II. WIN/LOSS MATRICES FOR THE 3 PERFORMANCE CLUSTERS FOR THE DIFFERENT STATES PREDICTED USING IR SENSORS.

State	Low-perf. cluster				Mid-perf. cluster				High-perf. cluster			
Agreeableness	0	0	0	0	0	1	1	1	0	6	7	8
	1	0	0	0	13	0	4	3	7	0	6	8
	3	3	0	2	14	10	0	6	6	5	0	6
	3	3	0	0	14	10	5	0	5	3	5	0
Conscientiousness	0	0	0	0	0	0	0	0	0	8	9	10
	1	0	0	1	14	0	2	1	9	0	6	7
	1	1	0	1	15	12	0	6	9	8	0	5
	2	1	1	0	15	13	4	0	8	6	5	0
Creativity	0	0	0	0	0	1	1	1	0	13	11	11
	4	0	2	2	15	0	6	8	3	0	8	6
	5	3	0	0	15	9	0	3	6	7	0	5
	7	5	6	0	15	7	5	0	6	8	4	0
Emotional Stability	0	0	0	0	0	2	2	2	0	3	4	3
	7	0	2	2	16	0	6	6	3	0	4	3
	9	5	0	1	16	11	0	6	2	3	0	3
	9	6	4	0	16	11	4	0	3	5	2	0
Extraversion	0	2	1	1	0	0	1	0	0	8	9	8
	7	0	0	2	20	0	9	9	4	0	5	7
	9	11	0	5	19	9	0	5	4	6	0	3
	10	10	3	0	20	9	5	0	5	4	5	0

State	Low-perf. cluster				Mid-perf. cluster				High-perf. cluster			
High PA	0	1	2	2	0	5	3	4	0	10	10	9
	5	0	2	2	15	0	4	5	1	0	5	6
	7	6	0	3	17	14	0	4	1	4	0	2
	7	6	2	0	16	13	6	0	2	3	4	0
High NA	0	0	0	1	0	1	0	0	0	2	0	0
	9	0	2	2	10	0	3	2	0	0	0	0
	9	6	0	3	12	8	0	3	1	2	0	0
	8	7	3	0	12	8	4	0	1	2	0	0
Low PA	0	0	0	0	0	0	0	0	0	4	3	3
	3	0	2	1	11	0	5	4	3	0	2	3
	3	2	0	0	13	7	0	2	4	4	0	2
	4	3	2	0	13	8	7	0	4	3	1	0
Low NA	0	0	0	0	0	0	0	0	0	11	7	9
	6	0	2	3	11	0	7	6	3	0	4	5
	8	4	0	2	12	4	0	4	7	8	0	5
	9	5	4	0	12	5	2	0	5	7	4	0

contributes to recognition improvements over a baseline. Furthermore the amount of improvement tends to increase with graphlet complexity. These results show the feasibility of the proposed approach, and hopefully encourage further research. We also find that distinct sensory channels play different roles for distinct target states: e.g. complex graphlets derived from IR have a large impact for Conscientiousness, while those derived from BT do not. The opposite trend is observed for Low NA. These findings support the intuition that the two channels capture different concrete behaviors: BT reflects proximity in a broadcast manner (i.e. many-to-one), IR is restricted to face-to-face (one-to-one) interactions.

ACKNOWLEDGMENT

S.T. and A.P. acknowledge support by grant PRIN 2009LNP494 (Statistical Relational Learning: Algorithms and Applications) from Italian Ministry of University and Research. The work of J.S. was partially supported by the FIRB S-PATTERNS project. The work of B.L. was supported by the Marie Curie-COFUND- 7th Framework PERSI project.

REFERENCES

[1] M. R. Barrick, M. K. Mount, and T. A. Judge, "Personality and Performance at the Beginning of the New Millennium: What Do We Know and Where Do We Go Next?" *International Journal of Selection and Assessment*, vol. 9, no. 1&2, pp. 9–30, 2001.

[2] T. A. Judge, D. Heller, and M. K. Mount, "Five-factor model of personality and job satisfaction: a meta-analysis." *Journal of Applied Psychology*, vol. 87, no. 3, pp. 530–541, 2002.

[3] S. Barsade and B. A., *The affective revolution in organizational behavior: The emergence of a paradigm*, 2nd ed. Londin:Lawrence Erlbaum Associates, Publishers, 2003, pp. 3–52.

[4] B. Staw and Y. Cohen-Charash, "The dispositional approach to job satisfaction: more than a mirage, but not yet an oasis," *Journal of Organizational Behavior*, vol. 26, pp. 59–78, 2005.

[5] A. Isen and A. Labroo, *Some Ways in Which Positive Affect Facilitates Decision Making and Judgment*, 2003, pp. 365–93.

[6] P. Salovery, J. Mayer, and D. Rosenhan, "Mood and helping: Mood as a motivator of helping and helping as regulator of moodl," *Review of Personality and Social Psychology*, vol. 12, pp. 215–237, 1991.

[7] W. Fleeson, "Toward a Structure- and Pricess-Integrated View of Personality: Traits as Density Distributions of States," *Journal of Personality and Social Psychology*, vol. 80, no. 6, pp. 1011–1027, 2001.

[8] ——, "Situation-based contingencies underlying trait-content manifestation in behavior," *Journal of Personality*, vol. 75, no. 4, pp. 825–861, 2007.

[9] E. Hatfield, J. Cacioppo, and R. Rapson, "Emotional Contagion," *Current Directions in Psychological Science*, p. 9699, 1993.

[10] J. H. Fowler and N. A. Christakis, "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study," *British Medical Journal*, vol. 337, 2008.

[11] S. G. Barsade, "The Ripple Effect: Emotional Contagion and Its Influence on Group Behavior," *Administrative Science Quarterly*, vol. 47, pp. 644–675, 2002.

[12] A. Hill, D. Rand, M. Nowak, and N. Christakis, "Emotions as infectious diseases in a large social network: the sisa model," *Proceedings of Royal Society*, 2010.

[13] S. T. Moturu, I. Khayal, N. Aharony, W. Pan, and A. Pentland, "Using social sensing to understand the links between sleep, mood, and sociability," in *SocialCom/PASSAT*, 2011, pp. 208–214.

[14] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro, "Multimodal recognition of personality traits in social interactions," in *10th int. conf. on Multimodal interfaces*, 2008, pp. 53–60.

[15] B. Lepri, S. Ramanathan, K. Kalimeri, J. Staiano, F. Pianesi, and N. Sebe, "Connecting meeting behavior with extraversion - a systematic study," *T. Affective Computing*, vol. 3, no. 4, pp. 443–455, 2012.

[16] J. Staiano, B. Lepri, N. Aharony, F. Pianesi, N. Sebe, and A. Pentland, "Friends don't lie: inferring personality traits from social network structure," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2012, pp. 321–330.

[17] Y.-A. de Montjoye, J. Quoidbach, F. Robic, and A. Pentland, "Predicting personality using novel mobile phone-based metrics," in *SBP*, 2013, pp. 48–55.

[18] D. O. Olguın, P. A. Gloor, and A. S. Pentland, "Capturing individual and group behavior with wearable sensors," *Proceedings of the 2009 aaai spring symposium on human behavior modeling, SSS*, vol. 9, 2009.

[19] B. Lepri, J. Staiano, G. Rigato, K. Kalimeri, A. Finnerty, F. Pianesi, N. Sebe, and A. Pentland, "The sociometric badges corpus: A multilevel behavioral dataset for social behavior in complex organizations," in *SocialCom/PASSAT*, 2012, pp. 623–628.

[20] K. Ara, N. Kanehira, D. O. Olguín, B. N. Waber, T. Kim, A. Mohan, P. A. Gloor, R. Laubacher, D. Oster, A. Pentland, and K. Yano, "Sensible organizations: Changing our businesses and work styles through sensor data," *Journal of Information Processing*, vol. 16, pp. 1–12, 2008.

[21] D. O. Olgun, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland, "Sensible organizations: Technology and methodology for automatically measuring organizational behavior," *IEEE Trans on Sys, Man and Cybpart B: Cybernetics*, pp. 43–55, 2009.

[22] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the big-five personality domains," *Journal of Research in Personality*, vol. 37, pp. 504–528, 2003.

[23] N. Pržulj, "Protein-protein interactions: Making sense of networks via graph-theoretic modeling," *Bioessays*, vol. 33, no. 2, pp. 115–123, 2011.

[24] H. A. Soufiani and E. Airoldi, "Graphlet decomposition of a weighted network," *J. Mach. Learn. Res.*, vol. 22, pp. 54–63, 2012.

[25] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[26] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008.