Stresscapes: Validating Linkages between Place and Stress Expression on Social Media

Martin Sykora M.D.SYKORA@LBORO.AC.UK
Centre for Information Management, SBE, Loughborough University UK

Colin Robertson
Geography and Environmental Studies, Wilfried Laurier University CANADA
Ketan Shankardass
Psychology Department, Wilfried Laurier University CANADA
Rob Feick
School of Planning, University of Waterloo CANADA
Krystelle Shaughnessy
Psychology Department, University of Ottawa CANADA
Becca Coates
CIM, SBE, Loughborough University UK
Haydn Lawrence
Geography and Environmental Studies, Wilfried Laurier University CANADA
Thomas W. Jackson
CIM, SBE, Loughborough University UK

Abstract

Understanding how individuals and groups perceive their surroundings and how different physical and social environments may influence their state-of-mind has intrigued re-searchers for some time. Much of this research has focused on investigating why certain natural and human-built places can engender specific emotive responses (e.g. fear, disgust, joy, etc.) and, by extension, how these responses can be considered in place-making activities such as urban planning and design. Developing a better understanding of the linkages between place and emotional state is challenging in part because both cognitive processes and the concept of place are complex, dynamic and multi-faceted and are mediated by a confluence of contextual, individual and social processes. There is evidence to suggest that social media data produced by individuals in situ and in near real-time may provide novel insights into the nature and dynamics of individuals' responses to their surroundings. The explosion of user-generated digital data and the sensorization of environments, especially in urban settings, provide opportunities to build knowledge of place and state-of-mind linkages that will inform the design and promotion of vibrant place-making by individuals and communities.

In this paper we present a novel study, to be undertaken this summer within the Greater Toronto area in Canada, with 140 recruited participants who are frequent, geo-tagging, Twitter users. The goal of the study will be to assess emotional, acute and chronic stress experienced in urban built-environments and as expressed during daily activities. An existing automated semantic natural language processing tool will be validated through this study, and it is hoped that the methodology developed can be extrapolated to other urban environments as well, with a second validation study already planned to take place next year in London, United Kingdom.

1. Introduction

In recent years automated processing of rich, geo-tagged, social media text streams, such as Tweets and Facebook status updates is receiving considerable attention in the literature. This is largely motivated by the insights and value that such datasets were shown to provide (Chew & Eysenbach, 2010; O'Connor et al., 2010; Tumasjan et al., 2010; Abel et al., 2012). Social-media streams, in general, al-

Proceedings of the 2^{nd} International Workshop on Mining Urban Data, Lille, France, 2015.

low for observing large numbers of spontaneous, real-time interactions and varied expression of opinion, which are often fleeting and private (Miller, 2011). Miller (2011) further points out that some social scientists now see an unprecedented opportunity to study human communication with various applications and contexts, which has been an obstacle up until recently. O'Connor et al. (2010) demonstrated how large-scale trends can be captured from Twitter messages, based on simple sentiment word frequency measures. The researchers evaluated and correlated their Twitter samples against several consumer confidence and political opinion surveys in order to validate their approach, and have pointed out the potential of social-media as a rudimentary yet powerful polling and survey methodology. In her position paper De Choudhury (2013) suggests that mental health studies would benefit from employing social media, as it provides an unbiased collection of an individuals language and behaviour, and Coppersmith et al. (2014) further highlight how social media enables large scale analyses, which has not been previously possible with traditional methods. Eichstaedt et al. (2015) propose a strong argument in favor of employing social media to study heart disease mortality based on psychological characteristics gleaned from Twitter language use. Especially negative emotional language and expressions of stress play an important role. They argue that traditional approaches that use household visits and phone surveys are costly and have limited spatial and temporal precision.

Motivated by this initial evidence we will be investigating emotional acute and chronic stress as expressed in geo-tagged, in-situ social media language. Our primary focus will be the connection between expressions of stress and the geography of urban built environments; applying geo-spatial analysis methods to define dynamic stress landscapes, or *stresscapes* that will help us to understand how stress varies from place-to-place and from time-to-time within urban centres. As Schwartz and Germaine (2014) rightly point out, studies concerning the combination of social media, identity performance, and place are still rare. Hence, we particularly seek to contribute to recent research related to the linkages between place and expressions of personal or social stress. Research on this topic has traditionally focused on the role of either individual or contextual factors; however, it is necessary to investigate the interplay between individuals and the nature of their immediate surroundings. Assessment of stress is normally overly general, which makes it hard to compare the experience of stress across individuals; whereas focusing on the emotional dimensions of the stress response offers a more specific measure for analysis. We will recontextualize social media expressions through spatial modelling and integration with contextual geospatial datasets describing participants' immediate surroundings. This will lead to new in-sights into how emotional stress is related to particular conditions (e.g. traffic congestion), place types and designs (e.g. public versus private places, high versus low density) and times (e.g. commuting rush hours) within urban communities.

As far as the authors are aware this is the first study of its kind, which will be looking at various forms of stress, linkages to urban environments, and validation of a computational social media analysis tool against 'real' experiences of acute and chronic stress, using already well established and validated measures from literature.

The remainder of the paper is organised as follows. Section 2 introduces some background and prior work on stress in urban environments and the computational tool for emotion based stress detection. Method details and overall validation study design are presented in section 3. Section 4 concludes the paper and suggestions for future work are made.

2. Background

Intensive acute and chronic psychological stress appears to play a causal role in the onset of multiple chronic disease outcomes, such as asthma and obesity (Shankardass et al., 2009; 2014), engendering significant costs related to economic productivity, and health and social service spending (Daar et al., 2007). A body of evidence suggests that the built environment shapes how we experience and respond to stress (Shankardass, 2012). However, there is a critical gap in our understanding of how our environments shape our experience of stressors (e.g., social disorder) and influence how we cope with our perceived stress because of the availability (or lack thereof) of resources, e.g., safe park space (Shankardass, 2012). There is a lack of place-based measures of stress to facilitate research on these interrelationships.

This study uses a *conceptual framework* recently proposed by Shankardass (2012), which builds on Pearlin's stress process heuristic (Pearlin, 1999), where sources of stress that are perceived as stressful can manifest emotion-al, behavioural and physiological responses (e.g., negative affect, smoking and endocrine activation, respectively). Two critical mediators of these responses are resource appraisal and coping behaviours, while the neighbourhood built environment can present stressors and offer resources that condition how we cope in space and time. This conceptual framework guides our hypotheses about which confounders and moderators ought to be considered in building a prediction model of emotional stress on stress-related endocrine activation. These include personality differences, such as trait anxiety and pessimism (Chang, 2002) - *which may confound the relationship* - and low self-esteem (Dumont & Provost, 1999) - *which may increase the effect of per-*

ceived stress on chronic endocrine activation, as well as low social support (ibid.) - *which may increase the effect of perceived stress on chronic endocrine activation*, and sex and gender (Baum & Grunberg, 1991) with hard-to-predict moderating effects on the relationship. Chronic endocrine activation may be more likely where individuals adopt coping styles that do not effectively deal with stressors (e.g., avoidance coping, rather than approach coping or problem-oriented coping.

Taking all this into account, our overall goal is to further develop and validate an ontology of emotional stress (based on presence of negative and the lack of positive affect) that will facilitate measurement through semantic analysis of geo-tagged Twitter posts (Sykora et al., 2013) and assess the predictive validity of perceived psychological stress.

2.1. Detection of Stress from Tweets

There are numerous systems for effective, efficient and accurate sentiment and emotion detection from language. A broader overview of the various approaches is available in Thelwall et al. (2012). One of the popular techniques is based on the use of words and phrase dictionaries with known associated sentiment polarities or emotion categories; however, these dictionaries, although sometimes combined and semi-automatically generated for better cross-domain performance, are relatively flat and lack semantic expressivity. Even more recently Eichstaedt et al. (2015) still used a combination of simple dictionaries to perform their automated tweet analysis.

In this work we employ an ontology based approach, which is essentially a map of words and phrases with a much richer semantic representation than simple dictionaries. The system we will use is called EMOTIVE and is based on (1) a custom Natural Language Processing (NLP) pipeline, which parses tweets and classifies parts-of-speech tags, and (2) an ontology, in which emotions, related phrases and terms (including a wide set of intensifiers, conjunctions, negators, interjections), and linguistic analysis rules are represented and matched against (Sykora et al., 2013). EMOTIVE automatically detects expressions of eight well recognised and fine-grained emotions in sparse texts (e.g. Tweets). The system discovers the following range of emotions; anger, disgust, fear, happiness, sadness, surprise (also known as Ekman's basic emotions - Ekman and Davidson, (1994)), and confusion and shame, but at the same time differentiates emotions by strength (also known as activation level, e.g. fear - 'uneasy', 'fearful', 'petrified'). An evaluation of the system against other benchmarks performed in Sykora et al. (2013) showed excellent results, with a very high f-measure of .962. Given the rich representation of emotions and the ontology this is based on, we will link and extended this system into representing stress in its various shapes and forms, with the intention to validate this system against real experiences of stress (see next section for details on this validation study).

3. Methodology and Study Design

Emotional stress has been conceptualized in different ways, including in terms of negative affect and as a state of distress. Two criteria will be utilized as criteria for validation in this study, including;

- The single-item distress thermometer, which is a simple Likert scale shaped like a vertical thermometer that asks the subject to select a number corresponding to their level of distress (Zwahlen et al., 2008).

- The 10-item negative affect scale from the expanded version of the Positive and Negative Affect Schedule (PANAS-X) will also be used (Watson & Clark, 1994).

These measures will be framed using moment instructions, i.e., we will ask whether participants have experienced distress/negative affect "right now", that is, at the present moment. The aim will be to collect at least 10 measures of each during the two week follow-up (see section on overall design). An algorithm will be used to scan a series of discrete stress-related terms (still being compiled) in real-time for all participants and randomly trigger the study follow-up survey via SMS / text message. This survey will be triggered roughly at evenly-spaced time points across the two-week follow-up period, based on the rate of tweets by the study participants. In order to understand the context of Tweets, a series of questions will also be asked, to assess what activity mode the participant was in (e.g., work, play, commute, domestic, study) at the time of the Tweet, and whether and how the surrounding environment influenced the Tweet in any way. Participants will also be requested to automatically geo-tag their Tweets by default, for the duration of the study follow-up.

3.1. Recruitment

The study population will include long-term (>3 months at study entry) active (>4 posts per week) Twitter users who are free from anxiety disorders. The planned sample size is 140 participants, which was calculated based on Bland and Altman (1986) and scaled up by 40% for anticipated drop-outs. A pool of potential study participants (i.e. long term, active Twitter users) will be identified from a database of several million collected Tweets, geo-tagged in the Toronto area. The study will also be limited to participants who live and work in the greater Toronto area.

3.2. Overall Design

The study can be broken-up into three phases:

- a. Running enrollment of study participants (1 month, begins mid-May 2015)
- b. Follow-up period (2 months) each participant 2 weeks
- c. Study exit and hair sampling (running in parallel to b), and ultimately study takedown by mid-September.

Data about participants will be collected at study entry, specifically socio-demographic and psychological information. Information about the experience of emotional stress and relationship with place will be collected at approximately 10 time points during the follow-up period. Subsequently in the study exit participants will be asked to complete a checklist of potentially stressful events, in order to understand the influence of major life events during the follow-up period.

3.3. Assessment of Chronic Psychological Stress

The research team has also secured additional funding to augment this summer's study with a collection of hair samples for cortisol analysis in order to examine how our devised measure of stress predicts chronic activation and allostatic load (i.e. physiological dysfunction). Participants who agreed to this will provide a 0.95 cm hair sample (from the root) at *study exit*, which will be analysed using immunoassay analysis following a validated protocol (Gow et al., 2010). Because hair grows at a rate of approximately 1.25 cm per month, cortisol embedded in this sample length will reflect a retrospective record of approximately three prior weeks. Hair cortisol level will be considered an outcome in regression models from our measure of stress.

4. Conclusion and Future Work

There are several key benefits of our study. First, a place-based measure of physiological stress will significantly broaden the potential for research to examine how the neighbourhood environment affects human health and well-being. This could lead to studies that inform the design of neighbourhoods that facilitate stronger prevention and management of stress-related illnesses. Second, the final predictive validation model will create empirical evidence of the inter-relationship amongst emotional and psychological stress, endocrine activation and a range of demographic and psychological traits. The study described in this paper will be repeated, with lessons learned, next year in the city of London. It is hoped that this will strengthen the model and validation, and will also provide a culturally different built environment, with its own characteristics. We hope this will lend itself to some interesting analyses.

Acknowledgments

We are grateful for this work to be supported by an SSHRC (Social Sciences and Humanities Research Council) Partnership Development grant and partly by an internal Wilfrid Laurier University grant.

References

Abel, F., Hauff, C., Houben, G., Stronkman, R., and Tao, K. Semantics + filtering + search = twitcident exploring information in social web streams. In *Proceedings of the 23rd ACM International Conference on Hypertext and Social Media*, Milwaukee, USA, 2012.

Baum, A. and Grunberg, N. E. Gender, stress, and health. *Health Psychology*, 10(2):80–85, 1991.

Bland, J. M. and Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327(8476):307–310, 1986.

Chang, E. C. Optimismpessimism and stress appraisal: Testing a cognitive interactive model of psychological adjustment in adults. *Cognitive Therapy Research*, 26 (5):675–690, 2002.

Chew, C. and Eysenbach, G. Pandemics in the age of twitter: content analysis of tweets during the 2009 h1n1 outbreak. *PLOS One*, 5(11):e14118, 2010.

Choudhury, M. De. Role of social media in tackling challenges in mental health. In *Proceedings of the 2nd International Workshop on Socially-aware Multimedia*, pp. 49–52, New York, USA., 2013.

Coppersmith, G., Harman, C., and Dredze, M. Measuring post traumatic stress disorder in twitter. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, Ann Arbor, USA, 2014.

Daar, A. S., Singer, P. A., Persad, D. L., Pramming, S. K., Matthews, D. R., Beaglehole, R., ..., and Bell. Grand challenges in chronic non-communicable diseases. *Nature*, 450(7169):494–496, 2007.

Dumont, M. and Provost, M. A. Resilience in adolescents: Protective role of social support, coping strategies, self-esteem, and social activities on experience of stress and depression. *Journal of Youth Adolescence*, 28(3):343–363, 1999.

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., ..., and Seligman, M. E. Psychological language on twitter predicts county-level heart disease mortality. *Psychological Science*, 26 (2):159–169, 2015.

Ekman, P. and Davidson, R. J. (eds.). *The Nature of Emotion: Fundamental Questions*. Affective Science. Oxford University Press, 1994.

Gow, R., Thomson, S., Rieder, M., Uum, S. Van, and Koren, G. An assessment of cortisol analysis in hair and its clinical applications. *Forensic Science International*, 196(1):32–37, 2010.

Miller, G. Social scientists wade into the tweet stream. *Science*, 333(6051):1814–1815, 2011.

O'Connor, B., Balasubramanyan, R., Routledge, B., and Smith, N. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, Washington D.C., USA, 2010.

Pearlin, L. I. *The stress process revisited: Reflections on concepts and their interelationships*. Handbook on The Sociology of Mental Health. Plenum Press, New York, 1999.

Schwartz, R. and Germaine, R. H. The spatial self: Location-based identity performance on social media. *New Media & Society*, 2014. doi: 10.1177/1461444814531364.

Shankardass, K. Place-based stress and chronic disease: A systems view of environmental determinants. In *Rethinking Social Epidemiology*, pp. 113–136. Springer, Netherlands, 2012.

Shankardass, K., McConnell, R., Jerrett, M., Milam, J., Richardson, J., and Berhane, K. Parental stress increases the effect of traffic-related air pollution on childhood asthma incidence. In *Proceedings of the National Academy of Sciences*, volume 106, pp. 12406–12411, USA, 2009.

Shankardass, K., McConnell, R., Jerrett, M., Lam, C., Wolch, J., Milam, J., Gilliland, F., and Berhane, K. 2014. parental stress increases body mass index trajectory in preadolescents. *Pediatric Obesity*, 9(6):435–442, 2014.

Sykora, M., Jackson, T. W., O'Brien, A., and Elayan, S. Emotive ontology: Extracting fine-grained emotions from terse, informal messages. *IADIS International Journal on Computer Science and Information Systems*, 8(2):106–118, 2013.

Thelwall, M., Buckley, K., and Paltoglou, G. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.

Tumasjan, A., Sprenger, T. O., and Welpe, I. M. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, Washington D.C., USA, 2010.

Watson, D. and Clark, L. A. The panas-x. manual for the positive and negative affect schedule: expanded form. Technical report, University of Iowa, Iowa City, IA, USA, 1994. URL http://www2.psychology.uiowa.edu/faculty/clark/panas-x.pdf.

Zwahlen, D., Hagenbuch, N., Carley, M., Recklitis, C., and Buchi, S. Screening cancer patients' families with the distress thermometer (dt): a validation study. *Psycho-Oncology*, 17(10):959–966, 2008.