ADOC: Affective Dimension of Collaboration

From the call:

* *Describe and discuss the objectives and research questions you are addressing in your project and make the case for their importance to the humanities and/or social sciences. Discuss your proposed methodology for addressing these questions.*
* *Provide a clear and concise summary of a state of the art survey of the relevant field. The goal of the survey is to call attention to similar work in the area of study. If there are existing projects that are similar in nature to your project, please describe them and discuss how they relate to the proposed project. This environmental scan should make it clear that you are aware of similar work being done; it should explain how the proposed project contributes to and advances the field.*
* *Discuss how the project addresses the overarching goals of the T-AP Digging into Data Challenge as described in this RFP, including the added value to the research endeavour of the proposed transnational partnership.*
* *Describe in detail the data chosen for the project. Demonstrate appropriate authorisation related to intellectual property or privacy issues associated with the data needed for the project. In the Letters of Commitment Section below, provide letters from the data guardians, indicating permissions to use the data, if appropriate.*
* *Provide a concise history of the project, including information about preliminary research or planning, financial support and/or in-kind contributions already received, and resources or research facilities available.*
* *Describe how your research project will assist in the training of students and newer-to-the-field researchers on your team.*

In this research project, we address the question of the nature of human collaboration using large-scale digital data traces from online collaborative networks (OCNs). Our goal is to provide new theoretical insights into what makes human online collaborations succeed, and, complementarily, what makes them fail. For example, we will look at *inclusivity:* what aspects of the inclusion or exclusion of contributors in an online social network help to promote or impede efficient collaborative efforts? We will also study *interaction:* how can communication be improved so that important social relationships are maintained, and detrimental ones are extinguished? Finally, we will study *identity:* what social structures are needed to form and to maintain strong and effective online collaborative networks?

We propose that the answer to these and other questions lies in the online social coding network. In the modern economy, a vast amount of human effort is currently expended online. While much of this effort may be viewed as mainly directed towards social goals (i.e. a teen’s selfies sent to friends), a large and growing community is using online social networks such as GitHub to collaborate on work projects. In fact, a large amount of modern business is done online, with little or no personal contact.

While it may be intuitively appealing to think of OCNs as meritocracies in which each member’s contributions are evaluated based on their objective strengths, in fact it is the case that the social relationships between members are a crucial element behind the functioning of the network. In this project, we will analyze data from OCNs by framing it as a process of *affect control.*  Our proposal is based on the theoretical idea that strong and persistent ties in human networks are primarily *relational,* rather than *transactional* (Lawler *et al.,* 2009). Transactional ties are those in which a participant is replaceable (e.g. your transactions with a bus driver). They are short-lived and brittle, and each participant must form individually rational ideas about how best to optimize the relationship. Relational ties, on the other hand, are based on affective (socio-emotional) understandings of other people. They tend to be much stronger and longer-lasting. Humans often start an interaction at a transactional level, but, given enough time and repeated exposure, this often progresses to a relational interaction. Thus, over time, a transactional network becomes a relational one: a meritocracy is naturally replaced with a affective anarchy. Humans follow this trajectory because it simplifies the interaction: the affective (relational) ties provide heuristics that are “ecologically rational” (Martin, 2009) and that simplify hard processing tasks.

We have recently proposed a computational model of affective interaction based on symbolic interactionist sociological theory called *BayesACT* (Schroder, Hoey and Rogers, 2016). *BayesACT* is a model of both relational and transactional ties, but puts emphasis on the relational (socio-emotional) aspects as the motivational force behind human actions. *BayesACT* arises from the symbolic interactionist tradition in sociology and proposes that humans learn and maintain a set of shared cultural affective sentiments about people, objects, behaviors, and about the dynamics of interpersonal events. Humans use a simple affective mapping to appraise individuals, situations, and events as sentiments in a three dimensional vector space of evaluation (good vs. bad), potency (strong vs. weak) and activity (active vs. inactive). These mappings are shared by all humans in a cultural group, and are passed from one member to the other (are learned by repeated exposure to associations demonstrated by members of the group). The sentiments can be measured, and the culturally shared consistency has repeatedly been demonstrated to be extremely robust in large cross-cultural studies (Heise, 2009; Osgood, May and Miron, 1975). This consistency is a keystone of human intelligence. Humans use it to make predictions about what others will do, and to guide their own behavior. The shared sentiments, and the resulting affective ecosystem of vector mappings, encodes a set of social norms1 that, if followed by all members of a group, results in an equilibirium or social order (Goffman, 1963) which is optimal for the group as a whole, rather than for individual members. Humans living at the equilibrium “feel” good and want to stay there. *BayesACT* thus provides a single, simple and parsimonious normative prescription to maintain consistency in sentiments (shared appraisals of events in an affective vector space). Further, *BayesACT* combines this relational model with a probabilistic model of rational decision making arising from a modern artificial intelligence tradition. The model is a partially observable Markov decision process, or POMDP, which can capture the complexities of dynamic (temporal) decision sequences, and can be used to find optimal solutions to these complex decision problems. Thus, *BayesACT* is ideally suited as a theoretical basis on which to frame the connections between relational (socio-affective) ties in an OCN with the transactional (outcomes measures) nature of the task that a group is trying to solve (e.g. write a complex piece of software).

Our research project will create and learn a *BayesACT* model of online collaborative networks. The model will be learned from GitHub data automatically. It will then be used in simulation to (A) validate it by comparing to real data; and (B) make predictions about future developments, providing even stronger validation; and (C) make recommendations about how collaborations can be fostered, improved, and made more (or less) inclusive using social restructurings, role and task re-assignments, or changes in communication media. The process of building and validating the model will provide the answers to our key social science hypotheses about the nature of collaboration. For example, a *BayesACT* model of a specific OCN might indicate that one person is behaving towards another as a *bully (EPA: -2.61, 1.17, 1.38),* while another as a *victim (EPA:* *-1.33, -2.42, -1.61).* Although these identities are affectively consistent, they may not be collaboratively efficient. Our analysis would then frame the interaction using the theoretical grounding of *BayesACT,* and provide alternatives for actions as suggestions to improve the relational ties between the group members.

Thus, our proposal addresses the following three research questions

1. Are online collaboration networks structured primarily using relational ties? Or perhaps: what are the relational structures present in OCNs?
2. What aspects of the relational structures in OCNs promote or impede efficient collaboration in terms of actual project outcomes?
3. What are the socio-cultural barriers and facilitators for effective online collaborations, including issues surrounding inclusivity by race, gender, or personality type.

Abundance of behavioural data on open collaboration platforms such as Github provides new opportunities for testing our hypotheses. Github is primarily used for Open Source Software (OSS) development (which is an increasingly popular way to build software). In the open source domain, humans seldom interact in person since they are not co-located. Hence, all interactions are in digital form. Also, because it is open source development, all digital interaction data is openly available. Therefore, as researchers, we have access to the interaction and development data for hundreds of thousands of open source software projects, that are being developed by millions of developers from around the world.

Recent papers (all 2014, 2015) on social and affective processes on Github demonstrate interest in this topic, but most are entirely empiricist, lacking theoretical foundation. In contrast, we will use Bayesian affect control theory, which is very well founded in decade-old knowledge from sociology and psychology, to study the affective basis of platform-based open collaboration. It is possible to learn affective identities, roles, ties and dynamics from large databases of interactions (e.g. pull requests/chats) on social coding networks, and that this process will be guided/facilitated through the use of a consistent and theoretically well-grounded model of how relational ties form and are maintained based in sociological symbolic interactionism.

Work related to the affect theory of social exchange and relational cohesion theory has shown that network structures both enable and constrain the development of fruitful social relations. Structural power and dependence shape who is likely to exchange with whom. The structure and frequency of social exchange affect positive/negative emotions, perceived cohesion with others, and ultimately one’s behavioral commitment to collaborative efforts.

While we have a clear theoretical starting point in the literature, as an important line of work has studied and developed theory around the affective basis of collaboration, and this work has not yet been explicitly theoretically linked with affect control theory.

Our research will allow for

(1) a test of relational cohesion theory and, more broadly, the affect theory of social exchange in a natural setting which relaxes some of the scope conditions of the theory,

(2) a test of BayesACT in such a setting, and the further development of a network-based model of Bayesian affect control, and

(3) a theoretical bridge to be drawn between BayesACT and exchange theories, as we see how uncertainty is managed and beliefs formed about self and other in interaction, and learn how these mechanisms play into the affective processes involved in collaboration as described by Lawler, Thye, and Yoon in their previous work.

Thus, several high-impact theory papers could be generated through this work for publication in sociological journals.

Challenges/necessary steps or tasks:

Each of these concepts is theorized and operationalized in specific ways in the literature, and will need to be translated effectively from a laboratory setting to natural interactions on GitHub.

For example:

-commitment has been operationalized in the previous literature as staying in the relation when good alternatives exist (stay behavior), unilaterally providing token gifts to each other (i.e., gifts that have little or no extrinsic value, are given without knowledge whether the other is giving, and without explicit expectation of reciprocity), and contributing to a joint venture that constitutes a social dilemma (i.e., puts individual resources at risk).

-Develop BayesACT variant of extended (ABOS) ACT model

-make predictions about joining groups/splitting up from groups

-how to measure/estimate “setting EPAs” for GitHub projects?

(applicable to project as a whole:)

-how to figure out identity EPAs from GitHub data?

-how to figure out behaviour EPAs and/or labels analogously to Dave’s group simulator paper / Bales interaction process analysis

Benefits:

Lawler, Thye, and Yoon have recently proposed and tested a theory of social commitments, showing that affective ties solve two fundamental problems of social order in groups that cannot be resolved with cognitive ties alone: sustaining membership and achieving the joint gains of collaboration. They find that joint tasks generate a sense of shared responsibility leading members to attribute their emotions from task interactions to the group as an object.

Such an affective alignment predicts decision to join or abandon existing projects or to start new ones - when confronted with a certain project through exposure in one’s social network, there are two determinants of the decision to take part or not: the content/meaning of the project (“it would be really, really cool to be part of developing Linux”) and the anticipated interaction experiences (“people seem to be kind of bossy here, which I really don’t like”). Both could be captured in parallel by an extended ACT model that would include settings (i.e., the projects) as per Lynn Smnvith-Lovin’s 1980’s work.

The learned model will combine with mined identities in specific situations to produce useful decision support for social coders. It will provide guidance and help in the structuring of a social groups online to make more productive networks. It will identify strengths and flaws in networks and report these to group members. It will provide online assistance during interactions to make them more effective.

Data to be used:

The various types of data that we can get from open source projects in Github are as follows:

1) Issue reports created by both developers and users of the software.

2) The discussion on each issue carried out by a subset of developers trying to build a solution for the issue.

3) Which solutions were accepted and which were rejected.

4) The description and rationale for the solution from the source code changes and comments.

5) The quality of the solution (by examining if more bugs occur in after the solution has been implemented).

6) The developers who have been making changes to a particular part of the software.

7) The list of people who collaborate on a particular task.

8) The complete list of changes done by a specific developer to a software project (along with when it was made).

9) The contact details of the developers in case we need to reach out to them to clarify our findings.

10) (Possibly) the social media handles for the developers.

11) The complexity of the changes and other software engineering metrics that we can use as control variables to compare apples to apples.

History of the project:

Our project team is comprised of a sociologist (Rogers), a cognitive scientist (Schroder), a computer scientist (Hoey), and a software engineer (Naggapan). Our unique combination of skills and knowledge applied to these research goals will ensure our success blah blah blah

- mention Tobias and Kim and I have worked together already on the theoretical basis, and Mei has worked on the OCN aspects.

other funding –

* I can mention my NSERC Discovery grant that has a section on studying social networks, but in a different context (elder care/gerontology).
* Kim’s grant to develop the simulator – we can leverage this so we’re not building a “tool” in this grant (something we are not supposed to do)

Other funding we can leverage?